PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 28 March 15, 1942 Number 3
Copyright 1942 by the National Academy of Sciences

HYPOXANTHINE, A GROWTH SUBSTANCE FOR PHYCOMYCES

By William J. Robbins and Frederick Kavanagh

New York Botanical Garden and Department of Botany, Columbia University

Communicated January 23, 1942

In previous papers from this laboratory an unidentified growth substance for *Phycomyces* widely distributed in products of natural origin was described as factor Z_1.[1] Guanine was found to have the same effect as factor Z_1, but was not considered to be identical with it because the activity of the former was destroyed by treatment with nitrous acid while that of the latter was unaffected. Xanthine, as might be anticipated from the effect of nitrous acid on the activity of guanine, was found to be inactive. Choline, adenine, thymine, uracil and cytosine also were ineffective.[2] Further experiments have been carried out in an attempt to identify factor Z_1.

The procedure followed was that described in the previous paper.[2] The growth of *Phycomyces* during a period of 72 hours at 26°C. was observed in 25 ml. of a basal mineral-dextrose solution containing asparagine and thiamine to which the various compounds or preparations tested for factor Z_1 activity were added alone and in combination with a D_R fraction[3] prepared from white potato tubers. We have used also the 48-hour development in 10 ml. of solution, a modification which economizes time and material.

The following were tested for factor Z_1 activity: monomethyl guanosine, monoacetone guanosine, guanosine picrate, guanylic acid, sodium guanylate, isoguanine sulfate, sodium inosine, hypoxanthine, vicine, divicine sulfate, barium inosinate and 2-amino uric acid (all obtained through the courtesy of the Rockefeller Institute for Medical Research), hypoxanthine (prepared in our laboratory by treating adenine with nitrous acid), a concentrate of folic acid supplied by R. J. Williams and one obtained from W. H. Peterson, uric acid, amygdalin, allantoin, theobromine, theophyllin, theocin,[4] adenine, guanine and xanthine.

Guanine and both samples of hypoxanthine were active and their activity on an equimolecular basis was approximately the same (table 1). None of the other compounds or the two concentrates of folic acid was

effective, at least in amounts of the same order of magnitude as the effective quantities of guanine or hypoxanthine. Hypoxanthine[5] was isolated in crystalline form as the silver salt (Fig. 1) from a Ca fraction from white potato tubers which had been treated with nitrous acid and electrodialyzed and which showed factor Z_1 activity. From 60 ml. of the solution 15.5 mg. of the silver salt of hypoxanthine were obtained. The substance isolated from this salt was active. On this basis it was estimated that the factor Z_1 activity of the extract from which the hypoxanthine was isolated could be accounted for by its hypoxanthine content. It would appear, therefore, that the active substance present in the potato extract after treatment with nitrous acid was hypoxanthine.

TABLE 1

THE EFFECT OF GUANINE AND HYPOXANTHINE IN THE PRESENCE OF THIAMINE AND A D_R FRACTION FROM POTATO TUBERS. NOTICE THAT BOTH COMPOUNDS ARE ABOUT EQUALLY EFFECTIVE IN INCREASING THE BENEFICIAL ACTION OF THE D_R FRACTION

ADDITIONS TO 10 ML. OF BASAL SOLUTION PLUS THIAMINE	DRY WT. 2 CULTURES MG.	ADDITIONS TO 10 ML. OF BASAL SOLUTION PLUS THIAMINE	DRY WT. 2 CULTURES MG.
10 μg. guanine HCl plus D_R fraction	63	6.6 μg. hypoxanthine plus D_R fraction	64
1 μg. guanine HCl plus D_R fraction	61	0.66 μg. hypoxanthine plus D_R fraction	67
0.1 μg. guanine HCl plus D_R fraction	43	0.066 μg. hypoxanthine plus D_R fraction	42
None	0.5	D_R fraction	27

The active material present in the original potato extract also appeared to be hypoxanthine. Examination of an active Ca fraction showed little or no guanine or xanthine, some adenine and considerable hypoxanthine.[6] The quantity of hypoxanthine found in this extract was sufficient to account for the activity of the fraction. The hypoxanthine crystallized from the potato extract was active; the adenine was inactive. Furthermore, the Ca fraction was as active (or more active) after treatment with nitrous acid as before. This would be expected if the active material in the original potato extract were hypoxanthine. We are inclined to believe, therefore, that factor Z_1 in potato extract is hypoxanthine though in other materials it may be guanine or a mixture of the two.

Hydrolysis of guanosine, guanylic acid, sodium guanylate, guanosine picrate, monomethyl guanosine, sodium inosine, barium inosinate and monoacetone guanosine yielded active preparations probably because of the guanine or hypoxanthine formed in the hydrolysis. Both samples of folic acid also gave active material on hydrolysis, that from Peterson being

the more active. This indicates that folic acid probably contains guanine or hypoxanthine.

The small quantity of hypoxanthine which is effective (activity has been observed for 0.3 mμ mole) is evidence for its functioning as a growth substance. Furthermore, the inactivity of xanthine, adenine and other compounds closely related to guanine or hypoxanthine demonstrates a considerable degree of specificity for the action of these two compounds and suggests that they may function in an enzyme system as thiamine, riboflavin, adenine and nicotinic acid are known to function. The inactivity of xanthine shows that the replacement of the NH_2 radical or hydrogen in the second position on the purine ring with oxygen renders guanine or hypoxanthine inactive (table 2). The inactivity of adenine suggests that oxy-

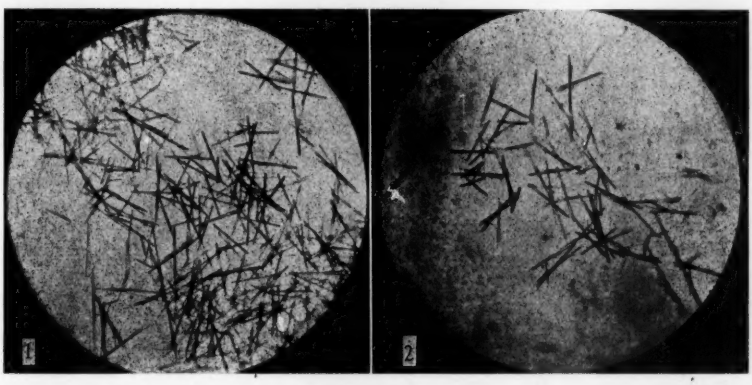


FIGURE 1

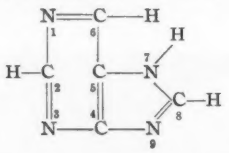
(1) Crystals of the silver salt of hypoxanthine prepared from pure hypoxanthine. (2) Crystals obtained with ammoniacal silver nitrate from potato extract which had been treated with nitrous acid and which showed factor Z_1 activity. Magnified 160 ×.

gen in the sixth position is important because the substitution of NH_2 for oxygen in hypoxanthine renders that compound inactive. The inactivity of 2-amino uric acid suggests that the hydrogen in the eighth position is important and may not be replaced by oxygen. The inactivity of guanylic acid, guanosine and similar compounds indicates that position 7 or 9[7] must be open, perhaps for combination with some other compound than the sugar which is present in guanosine. The inactivity of isoguanine emphasizes the importance of oxygen in the sixth position and hydrogen or NH_2 in the second position. It would appear that the active substance must have the eighth position and the seventh or ninth position open for substitution, oxygen in the sixth position and hydrogen or NH_2 in the second position.

We can only speculate as to whether hypoxanthine (or guanine) is active *per se*, or whether it, together with an unidentified substance, makes up a larger and active molecule, as, for example, thiazole and pyrimidine combine to form thiamine. Our observations have shown that factor Z_1 (hypoxanthine or guanine) is much more effective in the presence of a second unidentified substance, factor Z_2, than it is when used alone. Further information on the relation between factor Z_1 and factor Z_2 will have to wait the isolation of the latter substance.

TABLE 2

POSITION ON THE PURINE RING OF VARIOUS RADICALS FOR SOME OF THE COMPOUNDS TESTED FOR FACTOR Z_1 ACTIVITY



SUBSTANCE	POSITION ON PURINE RING							ACTIVITY
	1	2	3	6	7	8	9	
Guanine	H	NH_2	..	O	H	H	..	Active
Hypoxanthine	H	H	..	O	H	H	..	Active
Isoguanine	..	O	H	NH_2	H	H	..	Inactive
Xanthine	H	O	H	O	H	H	..	Inactive
Adenine	..	H	..	NH_2	H	H	..	Inactive
2-amino-uric acid	H	NH_2	..	O	H	O	H	Inactive
Theobromine	H	O	CH_3	O	CH_3	H	..	Inactive
Theophyllin	CH_3	O	CH_3	O	H	H	..	Inactive
Guanosine	H	NH_2	..	O	sugar	H	..	Inactive

[1] Robbins, W. J., *Am. Jour. Bot.*, **26**, 772–778 (1939); Robbins, W. J., *Bot. Gaz.*, **101**, 428–449 (1939); Robbins, W. J., *Am. Jour. Bot.*, **27**, 559–564 (1940); Robbins, W. J., and Hamner, K. C., *Bot. Gaz.*, **101**, 912–927 (1940); Robbins, W. J., *Ibid.*, **102**, 520–535 (1940).

[2] Robbins, W. J., and Kavanagh, F., *Proc. Nat. Acad. Sci.*, **28**, 4–8 (1942).

[3] The Ca fraction was the material adsorbed from an extract of potato tubers on charcoal and eluted with ammoniacal acetone. The D_R fraction was the filtrate from the charcoal treated extract.

[4] Theocin is a trade name for theophyllin. Our preparation of theocin was synthetic.

[5] The crystals of the silver salt of hypoxanthine were prepared as follows: A Ca fraction from potato tubers was treated with nitrous acid and electrodialyzed. The cathode fraction was evaporated to small volume and the purines precipitated with ammoniacal silver nitrate. The precipitate was dissolved in dilute hot nitric acid. The solution was cooled, the precipitate removed and redissolved in hot dilute nitric acid. This procedure was repeated five times.

[6] The precipitate formed from 500 ml. of a Ca fraction by ammoniacal $AgNO_3$ was freed from silver by treatment with HCl. Picric acid was added and the precipitate filtered off. Ammoniacal $AgNO_3$ was added to the filtrate and the precipitate was dis-

solved in hot dilute HNO_3. The precipitation with $AgNO_3$ and solution in HNO_3 was repeated five times. The pale yellow solution was treated with a small amount of Norit A and filtered. From the filtrate 15 mg. of crystals were obtained identical in appearance with the silver salt of hypoxanthine.

The precipitate of picrates was dissolved in dilute NH_4OH and treated with $AgNO_3$. The precipitate was filtered off, washed, decomposed with HCl to remove the silver and filtered. The filtrate was made slightly alkaline with NH_4OH and allowed to stand for 24 hours. No guanine precipitated. It was treated with nitrous acid to convert any guanine and adenine to xanthine and hypoxanthine, respectively. Silver nitrate was added and the precipitate examined. It appeared to consist of the silver salt of hypoxanthine. In the examination of the picrate precipitate volumes were kept to less than 20 ml.

[7] There appears to be some uncertainty as to whether the sugar in guanosine is attached to the purine ring in the seventh or the ninth position. It is usually represented in the seventh position.

SIZE GENES OF MICE

BY W. E. CASTLE

DEPARTMENT OF GENETICS, UNIVERSITY OF CALIFORNIA

Communicated January 21, 1942

In previous papers[1] it has been shown that several mutant genes of the house mouse, in addition to other effects which they exert, act also as size genes, either increasing or decreasing general body size.

The present paper contains a report on a study of the effects of two coat color genes, pink eye$_2$ (p_2) and lethal yellow (A^y) on body size, separately and in association with each other, but always on a background of homozygous brown (bb). In a previous paper it was shown that the influence of p_2 on body size is increased by association with brown. The same is probably true of A^y also, although qualitatively its action is the opposite of that of p_2, since it increases rather than decreases body size.

In order to learn more about the interaction of the three mutant genes, p_2, A^y and b, crosses were arranged which would produce as litter mates mice of the four genetic classes (1) pink-eyed brown ($aa\ bb\ p_2p_2$), (2) pink-eyed yellow ($A^y\ a\ bb\ p_2p_2$), (3) brown ($aa\ bb\ P_2\ p_2$) and (4) yellow ($A^y\ a\ bb\ P_2\ p_2$).

All four groups would be homozygous for b. Groups (1) and (2) would be homozygous for p_2 also, but groups (3) and (4) heterozygous for p_2. Groups (1) and (3) would be free from the A^y gene, but (2) and (4) would contain it. All possible combinations of p_2 and A^y would thus be obtained associated uniformly with bb. The symbol for brown will accordingly be omitted from the genetic formulae hereafter.

TABLE 1

AVERAGE BODY SIZE OF MICE OF FOUR DIFFERENT GENOTYPES BORN AS LITTER MATES AND REARED TOGETHER, SEXES APART

GENOTYPE	NO.	WEIGHT	REL. SIZE	BODY	REL. SIZE	TAIL	REL. SIZE
(1) ♂ *aa* p_2 p_2	83	29.41 ± 0.28	91.9	94.76 ± 0.22	98.4	81.40 ± 0.26	98.1
(2) ♂ $A^v a$ p_2 p_2	88	35.58 ± 0.51	111.2	96.87 ± 0.25	100.6	82.47 ± 0.28	99.3
(3) ♂ *aa* P_2 p_2	58	31.98 ± 0.32	100	96.26 ± 0.31	100	83.00 ± 0.41	100
(4) ♂ $A^v a$ P_2 p_2	68	37.40 ± 0.56	116.9	98.12 ± 0.26	101.9	85.40 ± 0.26	102.8
(1) ♀ *aa* p_2 p_2	72	24.60 ± 0.23	100.9	90.94 ± 0.33	99.8	79.19 ± 0.29	97.1
(2) ♀ $A^v a$ p_2 p_2	79	30.31 ± 0.39	124.3	93.05 ± 0.28	102.1	80.31 ± 0.29	98.5
(3) ♀ *aa* P_2 p_2	33	24.37 ± 0.38	100	91.12 ± 0.25	100	81.50 ± 0.52	100
(4) ♀ $A^v a$ P_2 p_2	78	33.67 ± 0.58	138.1	95.21 ± 0.26	104.4	83.05 ± 0.30	101.9
All ♂ p_2p_2, $\frac{(1)+(2)}{2}$		32.49	93.6	95.81	98.5	81.93	97.3
All ♂ P_2p_2, $\frac{(3)+(4)}{2}$		34.69	100	97.19	100	84.20	100
All ♀ p_2p_2, $\frac{(1)+(2)}{2}$		27.45	94.5	91.99	98.7	79.75	96.9
All ♀ P_2p_2, $\frac{(3)+(4)}{2}$		29.02	100	93.16	100	82.27	100
Both sexes p_2p_2		...	94.0	...	98.6	...	97.1
All ♂ *aa*, $\frac{(1)+(3)}{2}$		30.69	100	95.51	100	82.20	100
All ♂ $A^v a$, $\frac{(2)+(4)}{2}$		36.49	118.9	97.49	102.1	83.93	102.1
All ♀ *aa*, $\frac{(1)+(3)}{2}$		24.48	100	91.03	100	80.34	100
All ♀ $A^v a$, $\frac{(2)+(4)}{2}$		31.99	130.6	94.13	103.4	81.68	101.6
Both sexes $A^v a$		...	124.7	...	102.7	...	101.8

In this study, as in earlier ones, the young mice were separated as to sex at weaning, and kept until six months old, when they were weighed, chloroformed and measured as to body length and tail length.

In table 1 is shown for each sex separately the average size of the mice of each group as indicated by the three criteria, weight, body length and tail length.

Group (3) consisting of *aa* P_2 p_2 animals shows body size influenced presumably by the brown gene alone of those under consideration, since the A^y gene is absent and the p_2 gene present only as an unseen recessive, probably with little if any influence on body size. This group then should give us a standard for comparison of the effects of the genes p_2 and A^y in various combinations. Average body weight of males in group (3) is 31.98 g., average body length 96.26 mm., average tail length 83.00 mm. The data for group (3) females are less satisfactory because of insufficient numbers. For the present we may pass them by.

Group (1) consisting of *aa* p_2 p_2 animals shows body size influenced by both the brown gene and the pink eye gene in association. Average weight of males in this group is 29.41 g., average body length 94.76 mm., average tail length 81.40 mm. These are decreases (from the group (3) standard averages) of 8 per cent in weight, 1.5 per cent in body length and 2.0 per cent in tail length, which must be ascribed to the influence of p_2 now homozygous. Decrease is shown by females also of group (1) as regards body length (0.2 per cent) and tail length (2.9 per cent) but not as regards weight. Less importance attaches to these differences because of the insufficient number of females in group (3).

Group (2) consisting of A^y *a* p_2 p_2 individuals differs from group (1) only in the substitution of an A^y gene for *a*. A marked increase in body size results amounting in males to 20.9 per cent in weight, 2.2 per cent in body length and 1.3 per cent in tail length. In females the corresponding increases are 23.2 per cent in weight, 2.3 per cent in body length and 1.4 per cent in tail length. The A^y gene, as is well known, increases weight chiefly by leading to accumulation of fat especially in females, hence the large percentage of increase in weight.

Group (4), consisting of A^y *a* P_2 p_2 individuals, gives us by comparison with group (3) another opportunity to measure the influence of A^y in substitution for *a*. In both these groups the influence of p_2 is negligible, since it is present only as an unseen recessive. The differential factor is A^y.

Group (4) males show an increase over group (3) males of 16.9 per cent in weight, 1.9 per cent in body length and 2.8 per cent in tail length. For females the corresponding increases are 38.1 per cent in weight, 4.4 per cent in body length and 1.9 per cent in tail length. But these figures are less reliable because of the small number of females in group (3) as already noted.

The relative body size of mice of each of the four groups, taking group (3) animals as a standard of comparison, is shown in table 1. Also by combining the data on groups (1) and (2) we obtain average values for all animals homozygous for p_2, which may be compared with the averages for all animals heterozygous for p_2, obtained by combining the data for groups (3) and (4). These combination values are shown in the lower half of table 1.

By combining in a similar way the data for groups (1) and (3), we obtain values for all animals *free from the A^y gene* (*aa* animals) which may be compared with the values obtained by combining groups (2) and (4), which include all animals *possessing the A^y gene*. See the lowest section of table 1.

We thus obtain a comparison, based on all the available data, of the effectiveness of p_2 in decreasing body size and of A^y in increasing it, from which it appears that the influence of A^y is greater on weight and body length but not on tail length, which, however, is least valuable of the three criteria of body size.

The net percentage changes effected by the two genes are as follows:

	IN WEIGHT	IN BODY LENGTH	IN TAIL LENGTH
by $p_2\ p_2$	− 6.0%	−1.4%	−2.9%
by $A^y\ a$	+24.7%	+2.7%	+1.8%
Difference	+18.7%	+1.3%	−1.1%

These figures obtained by summarizing all the data furnished by the four groups agree well with those based on group (2) alone, in which both $p_2\ p_2$ and $A^y\ a$ were simultaneously operative, which are for weight a net percentage change of +17.7, for body length +1.3 and for tail length −1.1.

The results obtained in this experiment are similar both qualitatively and quantitatively to those obtained from earlier experiments. They show that the gene p_2 when homozygous decreases body size as judged by the three criteria, weight, body length and tail length. They show also that the gene A^y when heterozygous increases body size, notably in weight because of "adiposity," but also though in lesser degree in body length and tail length. When both influences are simultaneously in operation, the influence of A^y surpasses that of $p_2\ p_2$, much as regards weight, slightly as regards body length, but little or none as regards tail length.

Summary.—The mutant gene, pink-eye$_2$, decreases general body size in mice less than lethal yellow increases it, as regards both weight and body length, but not as regards tail length.

[1] See Castle, W. E., *Genetics*, **26**, 177–191 (1941).

THE OXIDATION OF 3,5-DIIODOTYROSINE TO THYROXINE*

BY TREAT B. JOHNSON AND LYNDON B. TEWKESBURY, JR.†

DEPARTMENT OF CHEMISTRY, YALE UNIVERSITY

Communicated February 2, 1942

A physiologically active protein product from which *thyroxine* (IV) can be isolated was prepared from milk casein by W. Ludwig and P. von Mutzenbecher[1] in 1939, by treatment of this protein with iodine under carefully controlled conditions. The experimental technique of these workers was repeated by C. R. Harington and Rosalind V. P. Rivers[2] in England, and the original findings of the German investigators were confirmed in every respect. There appears to be no doubt as to the correctness of the conclusions reached by these independent workers and it, therefore, becomes a matter of immediate chemical interest to consider any possible mechanism of reaction whereby such an important biochemical transformation can theoretically be brought about.

That iodine under proper experimental conditions is capable of affecting an oxidative coupling of two molecules of 3,5-diiodotyrosine to form a thyroxine molecule was established later by P. von Mutzenbecher,[3] who accomplished this change successfully by oxidation of 3,5-diiodotyrosine with hypoiodous acid formed during prolonged incubation of this α-amino-acid in alkaline solution at 37°. These experiments were repeated and the results reported by P. von Mutzenbecher[3] were confirmed by P. Block, Jr.,[4] in 1940. The original hypothesis advanced by C. R. Harington[2] that the biogenic synthesis of *thyroxine* probably results from the iodination of tyrosine functioning in the protein molecule through the intermediate stage of 3,5-diiodotyrosine must, therefore, receive careful consideration in the light of these most interesting results.

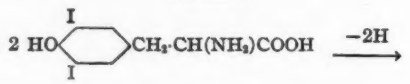
Biogenic Synthesis:

Tyrosine ⟶ 3,5-Diiodotyrosine ⟶ Thyroxine.

The authors have repeated the incubation experiments of P. von Mutzenbecher[3] using synthetical 3,5-diiodotyrosine of the highest purity, and have confirmed his conclusion that thyroxine is a product of reaction under specific experimental conditions. If the postulation be correct that thyroxine is produced here as a result of oxidation of the 3,5-diiodotyrosine by hypoiodous acid (HIO) formed from this α-amino acid in the alkaline solution, then the yield of thyroxine should theoretically vary according to the quantity of active hypoiodate present in the solution. The authors, therefore, modified P. von Mutzenbecher's technique[3] by adding hypoiodate to the α-aminoacid solution in molecular proportion with the result that the yield of thyroxine was slightly increased as postulated.

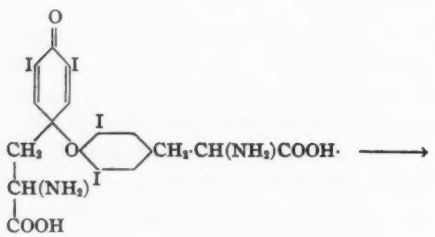
After applying several oxidation experiments with successful results, and after giving more attention to the nature of the secondary products of the reaction, the authors have been able to formulate a reaction mechanism which has not hitherto been presented to explain the formation of thyroxine (IV) from 3,5-diiodotyrosine (I). Our scheme of transformation is based on the results obtained in the fundamental studies of R. Pummerer[5] and coworkers dealing with the manner of oxidation of *o*- and *p*-substituted phenols in alkaline solution. We submit, therefore, the following graphical representation of this reaction (table 1) leading to the formation of thyroxine (IV) from 3,5-diiodotyrosine (I).

TABLE 1

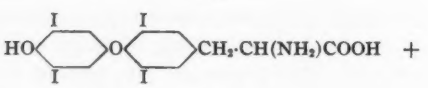


I

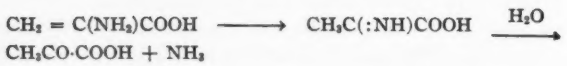
II



III



IV

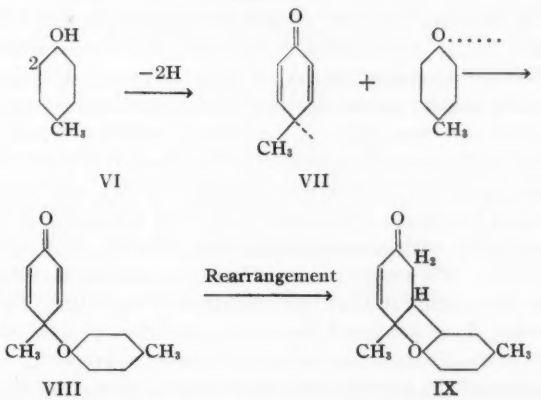


V

A comparison of the graphical representation of R. Pummerer's[5f] dehydrogenation of *p*-cresol (VI) by oxidation with $K_3Fe(CN)_6$ at 0° (table 2)

furnishes significant data which finds application in support of the hypothesis proposed in table 1.

TABLE 2



The chief product of this reaction is the tetrahydro-benzfurane structure expressed by formula (IX), and results apparently by rearrangement of the primary product of oxidation of *p*-cresol (VI) namely, the *p*-quinolether $C_{14}H_{14}O_2$ (VIII). In fact, according to R. Pummerer it is characteristic of *o*- and *p*-alkylated phenols to undergo dehydrogenation easily in alkaline media to form free radicals which dimerize to more or less stable aromatic quinolethers. Such free radicals are expressed by formulae (II) and (VII), and the dimerized products by formulae (III) and (VIII) in tables 1 and 2, respectively. That *o*- and *p*-alkylated phenols may react in such a tautomeric manner to give derivatives of dihydrobenzene was also shown earlier by K. von Auwers[6] in 1902, who synthesized the quinone structure (X), for example, by interaction of *p*-cresol with chloroform in alkaline solution; and also by T. Zincke[7] who demonstrated the production of the corresponding 1-methyl-1-trichlormethyl-4-ketodihydrobenzene (XI) by application of a Friedel and Craft's reaction with *p*-cresol and carbontetrachloride.

If the hypothetical, quinolether intermediate (III), represented as formed from 3,5-diiodotyrosine (I) (table 1), is compared structurally with the corresponding oxidation product of *p*-cresol (VIII) (table 2), it may be seen that the internuclear condensation or rearrangement, which takes place in the formation of the tetrahydro-benzofurane (IX) from the *p*-quinolether (VIII), is precluded in the case of the corresponding hypothetical 3,5-diiodotyrosine derivative (III). The substitution of iodine atoms in both nuclear positions *ortho* to the ether-bound oxygen atom in (III) prevents any possibility of cyclization of this compound. Hence, stabilization of the system expressed by structure (III) must follow one of two courses namely, (1) molecular dissociation with loss of one alanine side chain and formation of thyroxine (IV) and iminopyruvic acid or, (2) hydrolysis of (III) with production of serine, $HOCH_2CH(NH_2)COOH$ and thyroxine (IV). The results obtained by experimentation to date lead the authors to the conclusion that the formation of thyroxine (IV) from 3,5-diiodotyrosine (I) as described above is an oxidation process, and that the correctness of the reaction-mechanism postulated is supported by the facts that *pyruvic acid*[8] (V) and ammonia have been identified by the authors as secondary products of the reaction. Thus far, the authors have been unable to detect the presence of the α-aminoacid-*serine* as a product of the oxidation reactions. A thorough search for other products of reaction is now in progress.

Summary.—It is a well-known fact that thyroxine and also 3,5-diiodotyrosine slowly undergo decomposition especially when exposed to sunlight, in weakly alkaline solution with formation of hypoiodous acid (HIO).[9] This halogen acid interacts slowly with 3,5-diiodotyrosine to form thyroxine,[3, 4] but thus far no mechanism of reaction has been approved to explain this biochemical change. The authors have repeated the work of previous investigators of this important change, and have found that not only thyroxine but also pyruvic acid and ammonia are formed as secondary products of reaction. The discovery of *pyruvic acid* (V) as a product of this oxidation reaction has made possible the postulation of a simple and plausible mechanism to explain the formation of the *thyroxine*. The study of this problem is being continued and a discussion of the experimental technique and analytical results will be published elsewhere.

* Work supported in part by a grant from the George E. Sheffield Research Fund of the Sheffield Scientific School of Yale University.

† Present address: 222 Summer Street, Boston, Massachusetts. Ph.D. from Yale University in June, 1941.

[1] Ludwig, W., and von Mutzenbecher, P., *Zeit. physiol. Chem.*, **258**, 195–211 (1939).

[2] Harington, C. R., and Rivers, R. V. P., *Nature*, **144**, 205 (1939).

[3] von Mutzenbecher, P., *Zeit. physiol. Chem.*, **261**, 253–256 (1939).

[4] Block, P., Jr., *Jour. Biol. Chem.*, **135**, 51–52 (1940).

[5] (*a*) Pummerer, R., and Frankfurter, F., *Ber.*, **47**, 1472–1493 (1914).
(*b*) Pummerer, R., and Cherbuliez, E., *Ibid.*, **47**, 2957–2973 (1914).
(*c*) Pummerer, R., *Ibid.*, **52**, 1403–1413 (1919).
(*d*) Pummerer, R., and Cherbuliez, E., *Ibid.*, **52**, 1414–1415 (1919).
(*e*) Pummerer, R., and Frankfurter, F., *Ibid.*, **52**, 1416–1420 (1919).
(*f*) Pummerer, R., Melamid, D., and Puttfarcken, H., *Ibid.*, **55**, 3116–3122 (1922).
(*g*) Pummerer, R., Puttfarcken, H., and Schopflocher, P., *Ibid.*, **58**, 1808–1820 (1925).
(*h*) Pummerer, R., and Rieche, A., *Ibid.*, **59**, 2161–2175 (1926).

[6] von Auwers, K., and Winternitz, F., *Ibid.*, **35**, 465–471 (1902); von Auwers, K., and Keil, G., *Ibid.*, **35**, 4207–4217 (1902).

[7] Zincke, T., and Suhl, R. S., *Ibid.*, **39**, 4148–4153 (1906).

[8] The pyruvic acid is easily identified by the successful application of A. Baeyer's reaction involving the action of *o*-nitrobenzaldehyde on pyruvic acid which leads to the production of *indigo* (Baeyer, A., *Ibid.*, **15**, 2856 (1882); for other applications of this indigo color test see Johnson, T. B., and Baudisch, O., *Jour. Amer. Chem. Soc.*, **43**, 2670–2675 (1921); *Ber.*, **55**, 18–21 (1922).

[9] Kendall and Osterberg, *Jour. Biol. Chem.*, **39**, 125–129 (1919).

COMPLEMENT FIXATION WITH SIMPLE SUBSTANCES CONTAINING TWO OR MORE HAPTENIC GROUPS

BY DAVID PRESSMAN, DAN H. CAMPBELL AND LINUS PAULING

GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated February 13, 1942

It has been shown that simple compounds containing two or more haptenic groups undergo typical precipitin reactions with the antibody to a protein antigen containing homologous haptenic groups.[1] We have investigated the question of whether or not complement fixation takes place in the formation of precipitates with these simple compounds and have obtained the results reported below.

The compounds studied are listed in table 1. Phenylarsonic acid was included, although it does not form a precipitate with antiserum, because it is known to combine with antibody and thus to inhibit the formation of precipitate by the antibody and an antigen containing two or more benzene arsonic acid groups.

Experimental.—The substances I to IV and the rabbit antiserum homologous to azo sheep serum made with arsanilic acid were prepared by methods described elsewhere.[1c] The complement of the antiserum was inactivated by heating at 57° for 30 minutes.

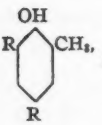
The amboceptor was prepared (by Mr. Carol Ikeda) by giving rabbits ten injections of 1 or 2 ml. of a 10% sheep cell suspension over a period of 3 weeks. The rabbits were bled on the seventh, eighth and ninth days after

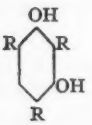
the last injection. Its complement was inactivated by heating at 57° for 30 minutes.

TABLE 1

COMPOUNDS INVESTIGATED

I. oxanilide-*p*,*p*′-diarsonic acid

II.  2-methyl-4,6-di-(*p*-azophenylarsonic acid)phenol

III.  , 1,3-dihydroxy,2,4,6-tri-(*p*-azophenylarsonic acid)benzene

IV. phenylarsonic acid

R represents —NN

Complement serum was pooled from the blood of four guinea pigs.

Complement fixation tests along with various controls were carried out by mixing saline solution, guinea pig complement (1/10 dilution) and haptene or antigen solution (20 μg./ml., pH 7.0) as shown in table 2. The mixtures were incubated for 30 minutes at 37° and the precipitates were

TABLE 2

PROTOCOL FOR COMPLEMENT FIXATION TESTS

TEST SOLUTION	COMPLEMENT 1/10 DILUTION ML.	ANTIAZOBENZENE *p*-ARSONIC ACID SERUM ML.	SALINE ML.	ANTIGEN OR HAPTENE (20 G./ML.) ML.	
1	1.5	0	3	0	Control on complement
2	1.5	1.5	1.5	0	Control on anticomplementary action of antiserum
3	1.5	0	1.5	1.5 I	Control on anticomplementary action of substances I to IV
4	1.5	0	1.5	1.5 II	
5	1.5	0	1.5	1.5 III	
6	1.5	0	1.5	1.5 IV	
7	1.5	1.5	0	1.5 I	Complement fixation tests of substances I to IV
8	1.5	1.5	0	1.5 II	
9	1.5	1.5	0	1.5 III	
10	1.5	1.5	0	1.5 IV	

removed from tubes 7, 8 and 9. To 0.5 ml. portions of each solution there were added 0.5 ml. portions of mixtures of equal volumes of 5% sheep cells and amboceptor of dilutions 1/25, 1/50, 1/100 and 1/200. The tubes were then incubated for 15 minutes at 37°. The results of the tests are shown in table 3.

TABLE 3

RESULTS OF COMPLEMENT FIXATION TESTS

TEST SOLUTION	Concentration of Amboceptor 1/25	1/50	1/100	1/200
1	4+	4+	+	−
2	4+	3+	+	−
3	4+	3+	+	−
4	4+	3+	2+	+
5	4+	3+	2+	+
6	4+	3+	2+	+
7	2+	−	−	
8	2+	−	−	−
9	2+	−	−	−
10	4+	3+	+	−

4 + means complete lysis
− means no lysis

Discussion and Summary.—The experimental results show that complement is removed from solution by the formation of a precipitate by a multihaptenic simple substance and its haptene-homologous antiserum. A simple univalent haptene, phenylarsonic acid, was found not to fix complement under the conditions investigated.

It is accordingly not necessary for fixation of complement that the antigen be a protein or other very complex molecule; but it is indicated that combination with antibody (as by haptene) without precipitation may not be sufficient for complement fixation.

[1] (*a*) Landsteiner, K., and van der Scheer, J., *Proc. Soc. Exp. Biol. and Med.*, **29**, 747 (1932); *Jour. Exp. Med.*, **56**, 399 (1932); (*b*) Pauling, L., Campbell, D. H., and Pressman, D., *Proc. Nat. Acad. Sci.*, **27**, 125 (1941); (*c*) Pauling, L., Pressman, D., Campbell, D. H., Ikeda, C., and Ikawa, M., *Jour. Am. Chem. Soc.*, to be published.

MAXIMAL SYLOW SUBGROUPS OF A GIVEN GROUP

By G. A. Miller

Department of Mathematics, University of Illinois

Communicated January 19, 1942

Suppose that a given solvable group G contains maximal Sylow subgroups of at least three different orders. Since G is solvable it contains an invariant subgroup of prime index p. This subgroup involves all the Sylow subgroups of G except those whose orders are powers of p. As the former Sylow subgroups are conjugate under this invariant subgroup as well as under G each of them is transformed into itself under G by p times as many operators as under this invariant subgroup. Hence none of these Sylow subgroups can be maximal under G unless it is of index p under G. That is, this invariant subgroup cannot contain maximal Sylow subgroups of G of more than one order. Hence *a solvable group cannot contain maximal Sylow subgroups of more than two different orders and if it contains maximal Sylow subgroups of two different orders at least one of them is invariant and of prime index under the group.*

If a group G contains maximal Sylow subgroups of exactly two different orders and two of them are invariant then they have only the identity in common and G is their direct product. That is, G is then the cyclic group of order pq, p and q being distinct prime numbers. If only one of these Sylow subgroups is invariant under G then it must be abelian since otherwise its commutator subgroup would be transformed into itself by an operator which would be prime to its order and G would involve maximal subgroups of more than two different orders according to the theorem at the close of the preceding paragraph. For similar reasons the given invariant maximal subgroup must then be of type 1^m and must admit an automorphism of prime order which does not transform any of its proper subgroups into itself. This condition is obviously sufficient as well as necessary.

From what precedes it results that a solvable group contains either no maximal Sylow subgroup or all its maximal Sylow subgroups are either of the same order or of two different orders. The last of these three cases is the simplest and was considered above. When all the maximal Sylow subgroups of a solvable group are of the same order and the group contains only one such subgroup this subgroup must be invariant and hence it must be of prime index under the group. It must therefore involve an operator whose order is equal to this index and transforms into itself at least one of the proper subgroups of the given invariant subgroup since otherwise the group would contain maximal Sylow subgroups of two different orders. The two abelian groups of order 12 are illustrations of solvable groups in which

all the maximal Sylow subgroups are of the same order and invariant, while in the dihedral and the dicyclic groups of this order the maximal Sylow subgroups are of the same order but non-invariant.

When every maximal subgroup of a given group, which may not be solvable, is a Sylow subgroup of G and G contains only two maximal subgroups then each of these subgroups is invariant under G and G is their direct product since they could not have any common operator besides the identity. Moreover, each of these two maximal subgroups is of prime order since otherwise G would involve a maximal subgroup which would not be a Sylow subgroup. That is, when every maximal subgroup of G is a Sylow subgroup and G contains only two maximal subgroups then G is the cyclic group of order pq, p and q being distinct prime numbers. When every maximal subgroup of G is a Sylow subgroup and G contains more than two maximal subgroups then it is not possible for all of these subgroups to be invariant under G since otherwise G would contain a maximal subgroup which would not be a Sylow subgroup. Hence at least one of these maximal subgroups appears in a complete set of conjugate subgroups of G involving more than one subgroup, and since all the Sylow subgroups of the same order contained in G appear in the same set of conjugates under G there is one and only one set of conjugate subgroups of G which involves this particular Sylow subgroup.

The given set of n conjugate subgroups is transformed under G according to a primitive permutation group since the subgroup composed of all the permutations of this group which omit a given letter is maximal. This primitive permutation group satisfies the condition that each of its subgroups composed of all its permutations which omit a given letter is of degree $n - 1$ and that all of its permutations besides the identity are of degree $n - 1$. For if two such subgroups had more than the identity in common then these common operators would be transformed into themselves by operators of each of these subgroups which are not common to the two subgroups since a proper subgroup of a prime power group is always transformed into itself by operators of this group which do not appear in this subgroup. It therefore would result that G would contain a maximal subgroup which would not be a Sylow subgroup since the proper subgroup composed of all the operators of G which would transform into themselves these common operators of the two given subgroups would either be a maximal subgroup of G or it could be extended by operators of G so as to obtain a maximal subgroup of G.

As this is impossible the given primitive group of degree n is of class $n - 1$ and hence it contains an invariant regular subgroup of order n according to a well-known theorem due to G. Frobenius.[1] This primitive permutation group is simply isomorphic with G since all the maximal subgroups of G are supposed to be Sylow subgroups of G. For the same reason the given in-

variant subgroup of order n is abelian and has an order which is a power of a prime number. It is of type 1^m since otherwise it would contain a characteristic proper subgroup and hence G would contain a maximal subgroup which would not be a Sylow subgroup. This proves the following theorem: *If every maximal subgroup of the group G is a Sylow subgroup then the order of G is divisible by two and only two distinct prime numbers and there is at least one invariant maximal subgroup in G. If G contains more than one invariant maximal subgroup it contains exactly two such subgroups and is the cyclic group of order pq, p and q being distinct prime numbers. If G contains only one invariant maximal subgroup this is abelian, of type 1^m, and of prime index under G.*

From this theorem it follows that when every maximal subgroup of a given group is a Sylow subgroup thereof then each of its Sylow subgroups is also a maximal subgroup. We proceed to consider the case when every Sylow subgroup of a given group is a maximal subgroup thereof. If such a group contains a maximal subgroup which is not also a Sylow subgroup this maximal subgroup cannot contain a Sylow subgroup of the group since this Sylow subgroup is supposed to be maximal. Its index under the group would therefore be divisible by each of the prime numbers which divide the order of the group. In particular, it could not be invariant under the group since a maximal invariant subgroup of a group is of prime index under the group. That is, if every Sylow subgroup of a group is a maximal subgroup of the group then the group cannot contain an invariant maximal subgroup unless it is also a Sylow subgroup, and if it is a Sylow subgroup it is of prime index under the group and comes under the theorem noted in the preceding paragraph.

Suppose that the group G contains maximal Sylow subgroups of order p^2, p being a prime number. If two of these subgroups have only the identity in common then every two of them have only the identity in common, and hence they are then transformed under G according to a primitive permutation group which contains a regular invariant subgroup whose order is equal to the number of these subgroups of order p^2 and hence its order would be prime to p. As the corresponding quotient group would be of order p^2 it would not be possible for every Sylow subgroup of G to be maximal. If two of the maximal subgroups of order p^2 would have a subgroup of order p in common this subgroup would be invariant under G and to a Sylow subgroup of the corresponding quotient group whose order would be prime to p there would correspond a subgroup of G which would not be a Sylow subgroup of G but would be either a maximal subgroup of G or could be extended to such a subgroup.

Hence it results that if every Sylow subgroup of G is maximal and G contains a Sylow subgroup of order p^2 then it contains only one such Sylow subgroup and hence is of order p^2q, p and q being distinct prime numbers,

and it contains an invariant subgroup of order p^2. This invariant subgroup is non-cyclic and none of its subgroups of order p is transformed into itself by an operator of order q. The tetrahedral group and a group of order 75 which involves the non-cyclic group of order 25 are instances of such a group. On the other hand when the given invariant subgroup is the non-cyclic group of order 9 operators of orders 2 and 3 in the group of isomorphisms would transform into itself a subgroup of order 3 contained in this invariant subgroup and hence not every Sylow subgroup would be maximal. It therefore results that *if every Sylow subgroup of a group is maximal and the group contains maximal subgroups which are not Sylow subgroups then the group must not only be insolvable but each of its Sylow subgroups must have an order which is at least the cube of a prime number.*

[1] Cf. Miller, Blichfeldt, Dickson, *Finite Groups*, second edition, p. 113 (1938).

AN INVERSE PROBLEM CONCERNING A CHAIN PROCESS

BY I. OPATOWSKI

INSTITUTE OF TECHNOLOGY, UNIVERSITY OF MINNESOTA

Communicated January 19, 1942

The processes in which one or more entities change their characteristics in a *successive manner* are *chain processes*. They have been studied in a deterministic and in a statistic manner.

In the *deterministic* treatment the entities involved in the process are grouped into a number of different kinds: {kind 0, kind 1, ..., kind n}. Let $N_i(t)$ be the amount of the kind i at the time t. If at the moment $t = 0$, in which the process started, all entities were of kind 0, then:

$$Y_0(0) = 1,\ Y_i(0) = 0 \text{ for } i \geqslant 1, \text{ where } Y_i(t) = N_i(t)/N_0(0). \quad (1)$$

If each kind i changes directly only into the kind $i + 1$ and if the amount of the kind i changed during dt into the kind $i + 1$ is proportional to $N_i(t)$, then:

$$dY_i/dt = k_iY_{i-1}(t) - k_{i+1}Y_i(t),\ (i = 0, 1, \ldots, n), \quad (2)$$

where k_i's are constants with $k_0 = 0$. If r_ρ constants k_i's are equal to each other, the symbol K_ρ will be used instead of k_i ($\rho = 1, \ldots, m$; $\sum_\rho r_\rho = n + 1$).

The system (2) as a mathematical expression of a chain process was introduced by Rutherford[1] to study the radioactive transformations, and

has been applied since then in many other fields. In chemical kinetics it represents a general type of successive reactions.[2] In the theory of biological effects of radiations it gives a relation between the number of killed microörganisms and the time of irradiation.[3] In the theory of photographic emulsions[4] it explains the formation of a latent image in relation to the time of exposure.

In the *statistic* treatment the process is a particular Markoff chain. All entities involved in the process are considered as forming a system which may take different states: {state 0, state 1,..., state n}. Let $p_{ij}(t)$ be the probability of being the system at the time $T + t$ in the state j, if it is at the time T in the state i. If the process is *homogeneous with respect to the time*, i.e., if $p_{ij}(t)$ is independent of T and if $p_{ij}(t)$ is continuous on the right at $t = 0$, then by a theorem of W. Doeblin[5] $p_{ij}(t)$ is necessarily a differentiable function when $t > 0$ and is differentiable on the right at $t = 0$. We consider a process of this type and restrict ourselves to the case in which an arbitrary state $i - 1$ may change during dt only into the state i, and the probability for this is $k_i dt$, where k_i (*an intensity*[6] of the process) is a constant. Then[7,8,6] (1) and (2) are satisfied by $p_{0,i}(t) \equiv Y_i(t)$ and also by the other probabilities $p_{ij}(t)$.

This statistic form of the process was first considered by H. Bateman[9] in connection with bombardment of a fluorescent screen by α particles. It occurs in general in radioactive disintegrations and was applied by W. Feller[10] to the growth of biological populations. Into the same form may be included also a chain process whose intensities depend on the time and are functions of the type: $k_i(t) = f(i)\varphi(t)$, because it reduces to a process with constant intensities by using a suitable time variable (of such a type is for instance the Pólya stochastic process[6]).

In many applications the only kinds or states of the changing system accessible to quantitative measurements are the final ones. For instance in killing microörganisms by irradiation, the various degrees of injury which the organisms undergo before dying, are usually not distinguishable with an accuracy necessary for a numerical counting, whereas dead organisms are clearly identified. The same situation occurs in radioactive transformations when the intermediate products are unstable or when their radiation energy is too small for measurements.[11] In such cases $Y_n(t)$ is the only Y_i-function which is obtained from the experiments and the problem arises to determine the laws governing the process, i.e., to evaluate the constants k_i's and the number n knowing $Y_n(t)$. This problem is the object of the present paper.[12] Essential tools for its solutions are the *Laplace transformation*, which Bateman[13] used to solve (2), and the *complete homogeneous symmetric functions*[14] defined by: $h_0 = 1$; $h_\nu(k_1, \ldots k_{n+1}) = \sum \Pi_{i=1}^{i=n+1} k_i^{\alpha_i}$, where $\sum$ is taken for all sets of non-negative integers α_i such that $\sum_{i=1}^{i=n+1} \alpha_i = \nu \geqslant 1$. Since this does not involve any compli-

cation we will take k_i's as complex numbers. We put also $k_i \neq 0$ for $1 \leqslant i \leqslant n$, because[6] if a $k_i = 0$, all $Y_j \equiv 0$ for $j \geqslant i$. For $|z| < |q_\rho^{-1}|$, the expansion:[14]

$$\Pi_{\rho=1}^{\rho=m}(1 - q_\rho z)^{-r_\rho} = \sum_{\nu=0}^{\infty} h_\nu(q_\rho|r_\rho;\ \rho = 1,\ldots, m)z^\nu, \qquad (3)$$

holds, where $h_\nu(q_\rho|r_\rho;\ \rho = 1,\ldots, m)$ means that each q_ρ is taken r_ρ times in calculating h_ν. We need an expansion in partial fractions:

$$\Pi_{\rho=1}^{\rho=m}(s + K_\rho)^{-r_\rho} = \sum_{\rho=1,\,\mu=0}^{\rho=m,\,\mu=r_\rho-1} C_{\rho\mu}(s + K_\rho)^{\mu-r_\rho}. \qquad (4)$$

Introducing here $s = z - K_\tau$, $K_\tau - K_\rho = q_\rho^{-1}$ for $\rho \neq \tau$, multiplying by z^{r_τ} and taking into account (3) we get: $\Pi_{\rho=1,\,\rho\neq\tau}^{\rho=m}(-q_\rho)^{r_\rho}(1 - q_\rho z)^{-r_\rho} = \sum_{\mu=0}^{\mu=r_\tau-1} C_{\tau\mu}z^\mu$ + terms in z of degree $\geqslant r_\tau$. Therefore, by (3):

$$C_{\tau\mu} = h_\mu \Pi_{\rho=1,\,\rho\neq\tau}^{\rho=m}(K_\rho - K_\tau)^{-r_\rho}$$

where h_μ stands for

$$h_\mu((K_\tau - K_\rho)^{-1}|r_\rho;\ \rho = 1,\ldots, m;\ \rho \neq \tau).$$

The Laplace transform[15] of $Y_n(t)$, i.e., $y_n(s) \equiv \mathfrak{L}\{Y_n(t), s\} \equiv \mathfrak{L}\{Y_n\}$ is:[13]

$$y_n(s) \equiv \int_0^\infty e^{-st} Y_n(t)dt = P_n \Pi_{\rho=1}^{\rho=m}(s + K_\rho)^{-r_\rho},$$

where $P_n = k_1 k_2 \ldots k_n$. (5)

Therefore by (4), since[15] $v!(s + K)^{-v-1} = \mathfrak{L}\{t^v e^{-Kt}\}$:

$$Y_n(t) = P_n \sum_{\tau=1,\,\mu=0}^{\tau=m,\,\mu=r_\tau-1} A_{\tau\mu} t^{r_\tau-\mu-1} e^{-K_\tau t},$$

where $A_{\tau\mu} = C_{\tau\mu}/(r_\tau - \mu - 1)!$ (6)

If all $r_\tau = 1$, then $A_{\tau\mu} = A_{\tau,0} = C_{\tau\mu} = C_{\tau,0} = \Pi_{i=1,\,i\neq\tau}^{i=n+1}(k_i - k_\tau)^{-1}$, by a formula of Bateman.[13] It is seen from (6) that[15]

$$\mathfrak{L}\{Y_n(t), s\} \text{ exists for } \Re s > -\Re K_\rho,\ \rho = 1,\ldots, m. \qquad (7)$$

$Y_n(t)$ may be written for $n \geqslant 1$ as a repeated convolution:[15]

$$Y_n(t) = P_n \exp(-k_1 t) * \exp(-k_2 t) * \ldots * \exp(-k_{n+1} t), \qquad (8)$$

which is equivalent to a repeated integral, because $e^{-kt} * F(t) = e^{-kt}\int_0^t e^{kt} F(t)dt$. It is seen from (6) or (8) that $Y_n(t)$ is a *symmetric function* of $k_1,\ldots, k_n$. Putting in (3) $q_\rho = -K_\rho$, $z = 1/s$, we get: $s^{n+1}\Pi_{\rho=1}^{\rho=m}(s + K_\rho)^{-r_\rho} = \sum_{\nu=0}^{\infty}(-1)^\nu h_\nu(K_\rho|r_\rho;\ \rho = 1,\ldots, m)s^{-\nu}$ for $|s| > |K_\rho|$. Therefore by (5): $y_n(s) = P_n \sum_{\nu=0}^{\infty}(-1)^\nu h_\nu(k_1,\ldots, k_{n+1}) \cdot s^{-\nu-n-1}$, where $h_\nu(k_1,\ldots, k_{n+1}) \equiv h_\nu(K_\rho|r_\rho;\ \rho = 1,\ldots, m)$, and by known theorems,[15] since $\mathfrak{L}\{t^\alpha\} = \alpha! s^{-\alpha-1}$:

$$Y_n(t) = P_n \sum_{\nu=0}^{\infty}(-1)^\nu h_\nu(k_1,\ldots, k_{n+1}) t^{\nu+n}/(\nu + n)!, \qquad (9)$$

which[16] by (6) or (8) is convergent for any finite t and k_i.

We give now three methods for the solution of our problem: the determination of the k_i's and of the n from $Y_n(t)$. The question what method is more convenient in a practical application depends on how $Y_n(t)$ is known. If it is known with a very good accuracy for small values of t method I may be chosen. Method III requires an accurate knowledge of $Y_n(t)$ for large values of t if n is large.

Method I. We have from (9): $d^\kappa Y_n(0)/dt^\kappa \equiv Y_n^{(\kappa)}(0) = 0$ for $\kappa < n$, $Y_n^{(n+\nu)}(0) = (-1)^\nu P_n h_\nu(k_1, \ldots, k_{n+1})$ for $\nu \geqslant 0$. Since $h_0 = 1$, n is obtained as the order of the first derivative of $Y_n(t)$ which is $\neq 0$ at $t = 0$. Besides this:

$$h_\nu(k_1, \ldots, k_{n+1}) = (-1)^\nu Y_n^{(n+\nu)}(0)/Y_n^{(n)}(0). \tag{10}$$

Therefore $k_1, \ldots, k_{n+1}$ *are determined as the roots of the equation*

$$x^{n+1} + a_1 x^n + \ldots + a_{n+1} = 0, \tag{11}$$

where $(-1)^i a_i$ are the elementary symmetric functions of $k_1, \ldots, k_{n+1}$ and may be calculated by means of the equations[14] $\sum_{u=0}^{u=i} h_{i-u} a_u = 0$, $a_0 = h_0 = 1$.

Method II. This method and the following one hold if $\Re k_i > 0$ for $i \leqslant n$ and $k_{n+1} = 0$, or if

$$\Re k_i > 0 \text{ for } i \leqslant n+1. \tag{12}$$

(These restrictions on k_i's are normally satisfied in all applications.) We assume first that (12) holds. Then by (6): $Y_n^{(\kappa)}(\infty) = 0$ for $\kappa \geqslant 0$ and

$$\int_0^\infty Y_n^{(\kappa)}(t)dt = -Y_n^{(\kappa-1)}(0) \text{ for } \kappa \geqslant 1. \tag{13}$$

Therefore $n+1$ is the smallest value of $\kappa > 0$, for which the integral (13) is $\neq 0$, and by (10):

$$h_\nu(k_1, \ldots, k_{n+1}) = (-1)^\nu \int_0^\infty Y_n^{(n+\nu+1)}(t)dt \Big/ \int_0^\infty Y_n^{(n+1)}(t)dt. \tag{14}$$

When h_ν's are known, k_i's are calculated as in the method I. It is possible to lower the degree of equation (11) by one. In fact, by (7), (12) and (5):

$$\mathfrak{L}\{Y_n(t), 0\} = y_n(0) = \int_0^\infty Y_n(t)dt = 1/k_{n+1}, \tag{15}$$

which gives k_{n+1}. Then[14]

$$h_\nu(k_1, \ldots, k_n) \equiv h_\nu(k_1, \ldots, k_{n+1}) - k_{n+1} h_{\nu-1}(k_1, \ldots, k_{n+1})$$

may be calculated by means of (14) or (10) for $\nu = 1, \ldots, n$, and $k_1, \ldots, k_n$ are determined as the roots of an equation of type (11) of degree n.

As soon as k_{n+1} is known from (15), n may be determined also from the formula

$$n = 1 + \lim_{t \to 0} [td \log_e |Y_n'(t) + k_{n+1} Y_n(t)|/dt], \tag{16}$$

which holds without any restriction on k_i's. To prove (16) write a power series for $Y_{n-1}(t)$ by changing in (9) n into $n-1$. From this it is easy to see that

$$\lim_{t \to 0}[tY'_{n-1}(t)/Y_{n-1}(t)] = n - 1. \tag{17}$$

On the other hand (2) gives for $i = n$: $\log|Y'_n(t) + k_{n+1}Y_n(t)| = \log|k_n| + \log|Y_{n-1}(t)|$. By differentiating this with respect to t and taking into account (17) we get (16).

We show now that the method remains substantially unchanged if $k_{n+1} = 0$ and if (12) holds only for $i \leqslant n$. Of course (16) is true and all the other formulas hold too if n is changed in them into $n-1$. But since by (2) $Y'_n(t) = k_nY_{n-1}(t)$, (13) holds also for $Y_n^{(\kappa)}$ if $\kappa \geqslant 2$. (15) gives now $\int_0^\infty Y_{n-1}(t)dt = 1/k_n$. Therefore (†) $\int_0^\infty Y'_n(t)dt = k_n\int_0^\infty Y_{n-1}(t)dt = 1$ and $n+1$ is the smallest value of $\kappa \geqslant 2$ for which the integral (13) is $\neq 0$. $\sum Y'_i(t) = 0$, by (2), because $k_{n+1} = 0$, therefore by (1) $\sum Y_i(t) = 1$. On the other hand from (†) and (1) we get $Y_n(\infty) - Y_n(0) = Y_n(\infty) = 1$ for $n > 0$, and we conclude that the system which was entirely in the state 0 at $t = 0$ is at $t = \infty$ entirely in the final state n.

Method III (*Method of Moments*). We assume first that (12) holds. Then by known theorems of the Laplace transformation:[15]

$$y_n^{(\nu)}(0) = (-1)^\nu\mathfrak{L}\{t^\nu Y_n(t), 0\} = (-1)^\nu\int_0^\infty t^\nu Y_n(t)dt \tag{18}$$

is convergent. On the other hand from (5): $k_{n+1}y_n(s) = \prod_{i=1}^{i=n+1}(1 + k_i^{-1}s)^{-1}$ which by (3) is $= \sum_{\nu=0}^{\infty}(-1)^\nu h_\nu(k_1^{-1}, \ldots, k_{n+1}^{-1})s^\nu$ for $|s| < |k_i^{-1}|$. Therefore: $k_{n+1}y_n^{(\nu)}(0) = (-1)^\nu\nu!h_\nu(k_1^{-1}, \ldots, k_{n+1}^{-1})$ and by (18):

$$\nu!h_\nu(k_1^{-1}, \ldots, k_{n+1}^{-1}) = k_{n+1}\int_0^\infty t^\nu Y_n(t)dt, \quad \nu = 0, 1, \ldots \tag{19}$$

This reduces the determination of the k_i's to the solution of an algebraic equation in the same way as in the method II.

If (12) holds only for $i \leqslant n$ and $k_{n+1} = 0$, all formulae of this method hold if we change in them n into $n-1$. Therefore, since now $Y'_n(t) = k_nY_{n-1}(t)$, (19) becomes: $h_\nu(k_1^{-1}, \ldots, k_n^{-1}) = (\nu!)^{-1}\int_0^\infty t^\nu Y'_n(t)dt$, which determines $k_1, \ldots, k_n$ in the same way as previously.[17]

[1] Rutherford, E., Chadwick, J., Ellis, C. D., *Radiations from Radioactive Substances*, Cambridge, 1930, pp. 11–12.

[2] See for instance Branch, G. E. K., Calvin, M., *Theory of Organic Chemistry*, New York, 1941, p. 357.

[3] Crowther, I. A., *Proc. Roy. Soc. London*, **B100**, 401 (1926); Opatowski, I., *Bull. Amer. Math. Soc.*, **47**, 704 (1941); **48**, 46 (1942).

[4] Cf. for instance Silberstein, L., *Jour. Optical Soc.* **31**, 343 (1941).

[5] Doeblin, W., *Bull. d. Sc. Math.*, **62**, 21–32 (1938); **64**, 35–37 (1940).

[6] Lundberg, O., *On Random Processes and Their Application to Sickness and Accident Statistics*, Uppsala, 1940, pp. 57–68.

[7] Kolmogoroff, A., *Math. Ann.*, **104**, 428–437 (1941).

[8] Feller, W., *Trans. Amer. Math. Soc.*, **48,** 513 (1940)

[9] Bateman, H., *Phil. Mag.*, 6 s., **20,** 704–707 (1910).

[10] Feller, W., *Acta Biotheor.*, **5,** Part I, 15–17, 19–21 (1939).

[11] Meyer, S., Schweidler, E., *Radioaktivitaet*, 2nd ed., Berlin-Leipzig, 1927, p. 32.

[12] If $k_i \neq k_j$, this problem is equivalent to that of fitting a set of numerical data to the sum of exponentials $\Sigma c_i \exp(-k_i t) \equiv Y_n(t)$. In the case in which c_i's are independent of k_i's various fitting methods are known: Aigner, F., Flamm, L., *Phys. Ztschr.*, **13,** 1151–1155 (1912); Walsh, J. W. T., *Proc. Phys. Soc. London*, **32,** 26–30 (1919); Levy, H., *Ibid.*, **34,** 108–113 (1922). Further references in Meyer-Schweidler[11] p. 61; also Bernstein, F., *Ztschr. ang. Math. u. Mech.*, **7,** 441–444 (1927) and Doetsch[15] p. 27–28. In our problem however c_i's are certain functions of k_i's, and these methods are not good.

[13] Bateman, H., *Proc. Phil. Soc. Cambridge*, **15,** 423–427 (1910). This paper remained unknown to some recent authors who gave less simple methods for the solution of (2): Arley, N., *Mat. Tidsskr.*, 49–51 (1939); Lundberg;[6] Feller.[8]

[14] Murnaghan, F. D., *Theory of Group Representations*, Baltimore, 1938,pp. 108, 112. These functions are called also homogeneous product sums; cf. Littlewood, D. E., *Theory of Group Characters*, Oxford, 1940, pp. 82, 83, 88 or, alef-functions of Wronski, *s. Enc. d. math. Wiss.*, **IB3b,** 465, 459.

[15] Doetsch, G., *Theorie der Laplace-Transformation*, Berlin, 1937, pp. 23, 148, 157, 43–48, 61–64 and Widder, D. V., *The Laplace Transform*, Princeton, 1941.

[16] Only the first term of this expansion was known, *s. Handb. d. Phys.*, **22,** 190.

[17] The author acknowledges helpful information received in preparing this paper from Profs. J. L. Doob, L. Silberstein, J. W. Tukey.

ON CONFIDENCE INTERVALS

BY EDWIN B. WILSON

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated December 23, 1941

In 1927 I called attention to the fact that many statements about probability are highly elliptical and illustrated the matter by the simple case of a point-binomial universe with unknown probability p and observed value p_0 in some sample.[1] Using the admittedly rough estimate of probability based on the standard deviation one ordinarily writes

$$p_0 - \lambda \sqrt{p_0 q_0/n} < p < p_0 + \lambda \sqrt{p_0 q_0/n}$$

and states that the probability that the true value p in the universe lies between the limits given may be had from a probability-integral table entered with a normal deviation of λ units. I urged that a better procedure would be to use for the standard deviation the value $\sqrt{pq/n}$ obtained from the unknown p of the universe which leads to

$$\frac{p_0 + t/2}{1+t} - \frac{\sqrt{p_0 q_0 t + t^2/4}}{1+t} < p < \frac{p_0 + t/2}{1+t} + \frac{\sqrt{p_0 q_0 t + t^2/4}}{1+t} \quad (1)$$

where $t = \lambda^2/n$, and to the statement that if p (unknown) should be outside of those limits the probability of observing the sample would be less than that obtained from entering the probability-integral table with a normal deviation of λ units. I was trying to emphasize that we know nothing about the value of p, which must have whatever value it did have in the universe from which the sample was drawn, but that we could set limits based on probability calculations such that if p lay between them the chance

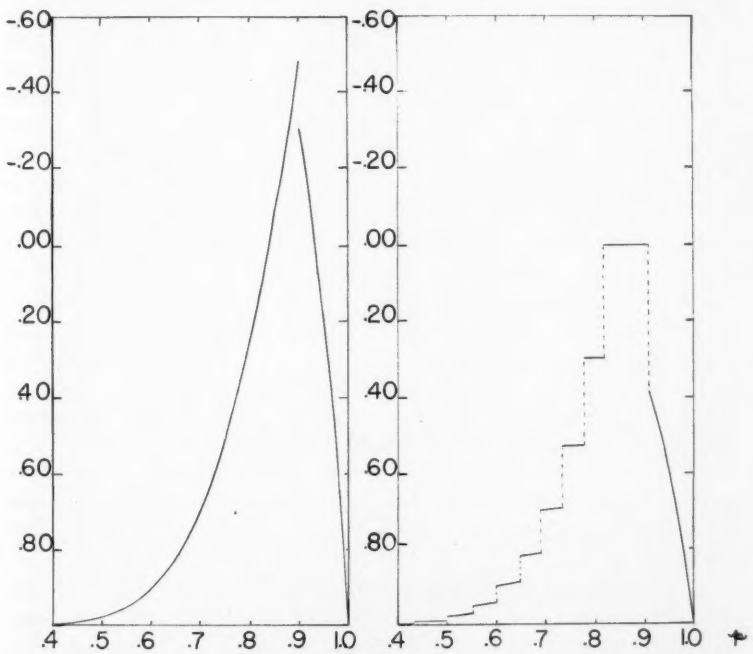


FIGURE 1

Confidence levels in terms of $1 - P = 1 - 2\epsilon$ for $n = 10$, $x = 9$, $p_0 = 0.9$ as functions of the unknown probability p on two assumptions:

1 (*a*), left: The no less extreme observation (Clopper and Pearson).

2 (*b*), right: The no more probable observation (Wilson).

of getting the particular observation or any less probable one would exceed some preassigned value P whereas if p lay outside them the chance of getting the observation or any less probable one would be less than P.

In 1934 Clopper and Pearson[2] introduced confidence intervals for what I take to be nearly the same notion that I was discussing. In their terminology, if 2ϵ be any positive number on which to base a certain level of con-

fidence in terms of probabilities, the interval (1) would have a confidence $1 - 2\epsilon = 1 - P$ as judged by the test I used. Their test was different; they asked the question: For what value p_1 of p less than p_0 will the value p_0 or a greater value arise with a probability ϵ and for what value p_2 of p greater than p_0 will the value of p_0 or a smaller value arise with the same probability ϵ? It may be noted that they divided the probability $P = 2\epsilon$ into two equal parts and asked two questions, as though one were thinking of deviations from a central tendency, whereas I had kept to a total probability P and had asked a single question in terms of samples no more probable than the one observed.[3] Furthermore Clopper and Pearson (and Rietz) introduce the notion of confidence belts, which I did not have in mind, and state: "We cannot therefore say that for any specified value of $x(= np_0)$ the probability that the confidence interval will include p is 0.95 or more. The probability must be associated with the whole belt, that is to say with the result of the continued application of a method of procedure to all values of x met with in our statistical experience"—a statement I should hesitate to make.

In figure 1 will be found for $n = 10$ the graph of the values of $1 - 2\epsilon = 1 - P$ against those of p for the case $p_0 = 0.9$ or $x = 9$ for (a) the Clopper-Pearson assumption that ϵ is the probability of all values 9 or less or of all values 9 or greater for x according as $p > 0.9$ or $p < 0.9$ and for (b) my own assumption that P is the probability of the sample of 9 or of any sample no more probable than it. The figure shows that on the Clopper-Pearson basis the curve is a continuous function of p except at the value $p = p_0 = 0.9$; it also shows that the shorter confidence intervals around $p = 0.9$ have confidence values (probabilities?) less than zero.[4] The figure shows also that on the basis of my definition involving the probability of all no more probable cases the curve of P has multiple discontinuities.

The question of generalization may be raised.[5] Let there be given two independent samples of n and n' from universes supposed to have the same unknown value p but furnishing from observation two values p_0 and p_0' in the two samples. If the numbers in the samples are large enough to justify estimates of probability based on values of χ^2, one may write

$$\chi^2 = \frac{(np_0 - np)^2}{npq} + \frac{(n'p_0' - n'p)^2}{n'pq}, \qquad (2)$$

$$\chi^2(p - p^2) = np_0^2 + n'p_0'^2 - 2(np_0 + n'p_0')p + (n + n')p^2,$$

which is a quadratic equation for the unknown value of p. Let

$$n + n' = N, \quad np_0^2 + n'p_0'^2 = N\bar{p}_0^2 + N\sigma_p^2, \quad t = \chi^2/N;$$

then

$$\frac{\bar{p}_0 + t/2}{1+t} - \frac{\sqrt{\bar{p}_0\bar{q}_0 t + t^2/4 - \sigma_p^2(1+t)}}{1+t} < p < \frac{\bar{p}_0 + t/2}{1+t} + \frac{\sqrt{\bar{p}_0\bar{q}_0 t + t^2/4 - \sigma_p^2(1+t)}}{1+t}. \quad (3)$$

This result is entirely similar to (1) except for the addition of the term $-\sigma_p^2(1+t)$ under the radical; it is, furthermore, entirely general for any number of samples drawn from different universes having the same p, provided $\bar{p}_0$ be the weighted mean p_0 for all samples and σ_p^2 the weighted variance of the observed values p_0.

The new term $-\sigma_p^2(1+t)$ under the radical, being negative, shows that for a given value of $t = \chi^2/N$ the range for p is the more restricted, the greater the scatter of the values p_0 observed in the different samples; indeed we must have[6]

$$\sigma_p^2 \leqq \frac{\bar{p}_0\bar{q}_0\, t + t^2/4}{1+t} \quad \text{or} \quad t \geqq \frac{2\sigma_p^2}{\sqrt{(p_0\bar{q}_0 - \sigma_p^2)^2 + \sigma_p^2} + \bar{p}_0\bar{q}_0 - \sigma_p^2} \quad (4)$$

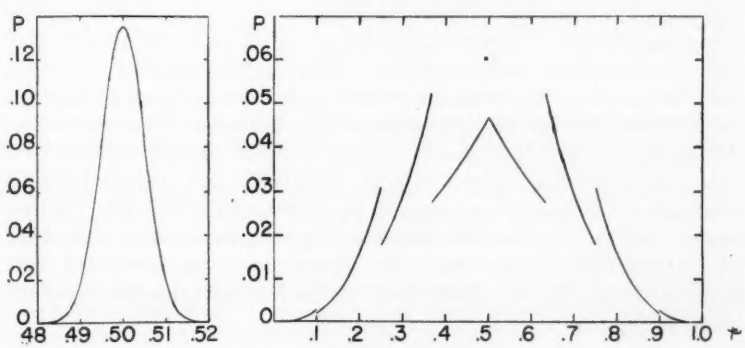


FIGURE 2

Confidence levels in terms of $P = 2\epsilon$ as functions of the unknown probability p on the hypothesis of the no more probable observation for two cases:
2 (*a*), left: $n = n' = 5000$, $p_0 = 0.49$, $p_0' = 0.51$.
2 (*b*), right: $n = n' = 3$, $p_0 = 1$, $p_0' = 0$.

to have any range for p. If we have k samples and enter a χ^2 table in the row for k degrees of freedom under any value of P to find χ^2 and hence $t = \chi^2/N$, N being the sum of the numbers in the k samples, we may compute from (3) the limits for p such that if p lies outside those limits the chance of getting the set of samples and all less probable sets is less than P; if p lies within the limits (2) the chance is greater than P but is in any case less

than the value of P corresponding to that value of t or χ^2 which makes the radical in (3) vanish or corresponds to the equality sign in (4) and thus shrinks the range of p to a point.[7] Figure 2*a* gives a plot of P against p for the case $n = n' = 5000$, $N = 10{,}000$, $p_0 = 0.49$, $p_0' = 0.51$, $\bar{p}_0 = 0.50$, $\sigma_p^2 = 0.0001$.

When the numbers are small the question of the test that replaces χ^2 has to be discussed. If we use the criterion of the total probability of the samples observed and of all samples of no greater probability, we may with patience discuss any particular case. For example, if $n = 3$, $p_0 = 1$, $n' = 3$, $p_0' = 0$, the joint chance of the observed samples is p^3q^3 whatever the unknown value of p. There are $(n + 1)(n' + 1) = 16$ possible pairs of samples. The graph of the probability P as a function of p is in figure 2*b*. The graph is, of course, discontinuous. It shows that there is no value of p for which the probability P exceeds 0.0625, and this for only the single value $p = 1/2$ which on *a priori* grounds would be infinitely rare. Values other than $p = 1/2$ between 0.37 and 0.63 give no values greater than $P = 0.047$. There are very short ranges for p near 0.36 and 0.64 which give $P > 0.05$. Should we desire confidence intervals[8] corresponding to $P = 0.05$ we should have to specify something like 0.360 to 0.366, 0.5, 0.634 to 0.640; should we desire confidence intervals corresponding to $P = 0.01$ we should have something like $0.17 < p < 0.83$; for $P = 0.03$ we should have 0.248 to 0.250, 0.302 to 0.366, 0.385 to 0.615, 0.634 to 0.698, 0.750 to 0.752. The exact method of R. A. Fisher, if he would apply it to a problem of this type (as is unlikely, because the approaches seem to belong to different universes of discourse) and if in this symmetrical case he would consider together both ends of his series, would give $P = 0.10$; there is no value of the unknown p which gives a probability as much as two-thirds of this amount. It thus appears that the exact method affords a different criterion from that of the total probability of the no more probable experience, just as both of these methods furnish different criteria from that based on the distribution of the difference $p_0 - p_0'$.

[1] Wilson, E. B., "Probable Inference, the Law of Succession, and Statistical Inference," *J. Amer. Statistical Association*, **22**, 209–212 (1927).

[2] Clopper, C. J., and Pearson, E. S., "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, **26**, 404–413 (1934). An account of this may be found in Rietz, H. L., "A Recent Advance in Statistical Inference," *Amer. Math. Monthly*, **45**, 149–158 (1938).

[3] There is a third notion not very different from mine or that of Clopper and Pearson—or rather of Neyman who slightly antedates them, at least with respect to continuous distributions—with which Fisher, R. A., on "Inverse Probability," *Proc. Camb. Phil. Soc.*, **26**, 528–533 (1930), associated the term fiducial limits and *according to him* limited to continuous distributions, see "The Fiducial Argument in Statistical Inference," *Ann. Eugenics*, **6**, 391–398 (1935).

[4] If I interpret Clopper and Pearson aright, one may logically take $2\epsilon = 1$, $\epsilon = 0.5$ and construct a confidence belt, quite as one constructs one for $2\epsilon = 0.05$. The belt is

of course thinner than for smaller values of ϵ but, still, has an appreciable area within which the probability that p will lie as "the result of the continued application of a method of procedure to all values of x met with in our statistical experience" is zero. As I do not see how this can well be, I incline to some doubt in respect to their (or Rietz's) statement. Moreover, for my test as shown in figure 1 (*b*) there is a finite interval for which $P = 1$ and the confidence is 0, whereas the probability that p lie in that interval could hardly be 0. Thus although confidence intervals are based on probabilities, it is not certain that probabilities are based on them. The difficulty of the negative values of probabilities $1 - 2\epsilon$ for p lying in intervals close around $p_0 = 0.9$ can be obviated by a modification of the definition whereby when $p \gtrsim p_0$ we use only a fraction θ of the maximum term M for $x = np_0$ and when $p < p_0$ the fraction $(1 - \theta)$ of M writing $X + \theta M = \epsilon$ and $(1 - \theta)M + Y = \epsilon$ where X is the sum of the terms up to but not including M, and Y is the sum of the terms beyond M, the value of θ being $1/2 + 1/2(Y - X)/M$.

[5] I shall not try to offer methods of generalizing the Clopper-Pearson test.

[6] While σ_p^2 may in extreme cases be nearly equal to $\bar{p}_0\bar{q}_0$ so that χ^2 may be near $2N\sigma_p = 2N(\bar{p}_0\bar{q}_0)^{1/2}$, the interesting range for σ_p^2 when N is large is a small multiple of $\bar{p}_0\bar{q}_0/N$ and thus σ_p^2 is small compared with $\bar{p}_0\bar{q}_0$, and approximately

$$\chi^2 \geqq \frac{\sigma_p^2}{\bar{p}_0\bar{q}_0/N}\left[1 + \frac{\sigma_p^2}{\bar{p}_0\bar{q}_0}\left(1 - \frac{1}{4\bar{p}_0\bar{q}_0}\right)\right].$$

The expression in front of the bracket is the usual expression for χ^2 taken from the marginal totals of a 2 by k table and the bracket is nearly 1. Throughout this paper as in that of 1927 I have carried out formal algebraic operations as though they were significant; it should, however, be borne in mind that the expression (2) for χ^2 is discontinuous and gives from a χ^2 table the probability of the given sample or set of samples and of all no more probable ones only to a certain degree of approximation even when the numbers are large—and thus the correction in the bracket may be illusory.

[7] It follows that in the case now under discussion the probability P which corresponds to a certain range of p such that if p be outside it the probability of so aberrant samples is less than P does not correspond to a confidence $1 - P$ that p lie within the interval. If P_0 be the probability which shrinks the interval to a point, we might perhaps say that $P_0 - P$ is the level of confidence that p lie within the interval and $1 - P_0$ the level of confidence that there be no value of p common to the two universes.

[8] As in the previous case, the chance if p lies outside the interval is less than the specified amount, but the chance if p lies within the interval is between the specified value and some upper limit P_0, leaving $1 - P_0$ as the level of confidence that there be no value of p common to the two samples. It may be repeated that the chances involved are not those that there actually is a common probability lying within or without certain limits but the chances that the observed samples and all no more probable should arise from such a hypothetical common value of p.

ON CONTINGENCY TABLES

BY EDWIN B. WILSON

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated January 8, 1942

In 1934, Yates,[1] acting on a suggestion from Fisher, developed the so-called exact method of treating 2×2 tables and proposed and discussed what has become known as the Yates' correction for the χ^2 expression used on such tables. Soon thereafter the method was inserted in a new edition of Fisher's Statistical Methods for Research Workers (Art. 21.02) and became rapidly and widely adopted. The method consists in considering the linear series of $c + d + 1$ tables

$$\frac{a-d}{c+d}\left|\frac{b+d}{0}\right., \ldots, \quad \frac{a-1}{c+1}\left|\frac{b+1}{d-1}\right., \quad \frac{a}{c}\left|\frac{b}{d}\right.,$$

$$\frac{a+1}{c-1}\left|\frac{b-1}{d+1}\right., \ldots, \quad \frac{a+c}{0}\left|\frac{b-c}{c+d}\right.,$$

where the convention has been adopted that $c + d$ is the smallest marginal total, and of computing the relative probabilities

$$\frac{(a+b)!(a+c)!(b+d)!(c+d)!}{(a+b+c+d)!}\left[\frac{1}{(a-d)!(b+d)!(c+d)!0!}, \ldots,\right.$$

$$\left.\frac{1}{a!b!c!d!}, \ldots, \frac{1}{(a+c)!(b-c)!0!(c+d)!}\right]$$

of those $c + d + 1$ tables. The properties of this series of probabilities, whose sum is 1, are therefore of great importance. Connected with them is the algebraic identity

$$\sum \frac{1}{a!b!c!d!} = \frac{(a+b+c+d)!}{(a+b)!(a+c)!(b+d)!(c+d)!}$$

where the summation extends over all the $c + d + 1$ terms of the series. The series is symmetrical when and only when two of the marginal totals on the same side of the table are equal.[2]

The mean value of a will be

$$\bar{a} = \frac{(a+b)!(a+c)!(b+d)!(c+d)!}{(a+b+c+d)!}\left[\sum \frac{a}{a!b!c!d!} = \right.$$

$$\left.\sum{}' \frac{1}{(a-1)!b!c!d!}\right]$$

The question is as to the value of the second sum $\sum'$ which is extended over all tables or terms. For those tables in which $a = 0$ the sum $\sum$ vanishes as does the sum $\sum'$ because $(-1)!$ is infinite. Hence in evaluating either sum one may reject the tables for which $a = 0$. Consider now the series of tables arising from $\begin{vmatrix} a-1 & b \\ c & d \end{vmatrix}$ which will have the same values of $(a-1)!b!c!d!$ as the given series. Then

$$\sum{}' \frac{1}{(a-1)!b!c!d!} = \frac{(a+b+c+d-1)!}{(a+b-1)!(a+c-1)!(b+d)!(c+d)!}$$

and hence $\bar{a} = \dfrac{(a+b)(a+c)}{a+b+c+d} = Np_1p_2,$

where $N = a+b+c+d$ and where $p_1 = (a+b)/N$, $p_2 = (a+c)/N$ are the proportions observed in the margins. The proof may be extended to cover the average of any (partial) factorial as:

$$\overline{a(a-1)..(a-p+1)} = \frac{(a+b+c+d-p)!(a+b)!(a+c)!(b+d)!(c+d)!}{(a+b+c+d)!(a+b-p)!(a+c-p)!(b+d)!(c+d)!}$$

From the values of the factorials the moments μ_i of the distribution of a may be obtained with routine algebra. Let

$$x = a - \bar{a} = (ad - bc)/N.$$

Then[3]

$$\mu_2(x) = \frac{N^2}{N-1} p_1p_2q_1q_2 = \frac{(Np_1q_1)(Np_2q_2)}{N-1}$$

$$\mu_3(x) = \frac{N^3}{(N-1)(N-2)} (q_1 - p_1)(q_2 - p_2)p_1p_2q_1q_2$$

$$\kappa = \frac{\mu_3}{\mu_2^{3/2}} \frac{\sqrt{N-1}}{N-2} \frac{(q_1 - p_1)(q_2 - p_2)}{\sqrt{p_1q_1p_2q_2}}$$

$$\mu_4 - 3\mu_2^2 = \frac{N^3p_1p_2q_1q_2}{(N-1)^2(N-2)(N-3)} [N^2(1 - 6p_1q_1 - 6p_2q_2 + 30p_1p_2q_1q_2) + N(6p_1q_1 + 6p_2q_2 - 36p_1p_2q_1q_2) - 1]$$

$$\beta_2 - 3 = \frac{N}{(N-2)(N-3)} \left[\left(1 - \frac{1}{N}\right)\left(\frac{1}{p_1q_1} - 6\right)\left(\frac{1}{p_2q_2} - 6\right) - 6 + \frac{1}{N}\left(1 - \frac{1}{N}\right) \frac{1}{p_1p_2q_1q_2} \right].$$

If we set

$$p = \frac{a}{a+b}, \quad p' = \frac{c}{c+d}, \quad p - p' = \frac{ad - bc}{(a+b)(c+d)} = \frac{Nx}{(a+b)(c+d)}$$

the characteristics of the distribution of $p - p'$, i.e., of the difference between the proportions observed in the first and second rows of the tables when the tables are restricted to the linear series used by Yates and Fisher, may be obtained at once, e.g.,

$$\mu_2(p - p') = \frac{N^4 p_1 p_2 q_1 q_2}{(N+1)(a+b)^2(c+d)^2} = \frac{1}{N-1}\frac{p_2 q_2}{p_1 q_1}.$$

If the two rows had arisen by independent sampling in two universes with probabilities p and p', respectively, and with numbers $n = a + b, n' = c + d$ in the two samples, there being $(n+1)(n'+1)$ different complexions for such pairs of samples, the value of μ_2 for the observed values of $p - p'$ would be $pq/n + p'q'/n'$ and if it should be assumed that p and p' were equal and had the value $(a + c)/N$ obtained from the marginal totals we should get $\mu_2(p - p') = p_2 q_2/(N p_1 q_1)$ which is slightly less than that obtained above. If it were assumed that the 2×2 table had arisen by sampling from a non-associated universe with probabilities $p_1 p_2$, $p_1 q_2$, $p_2 q_1$, $q_1 q_2$ for the four cells occupied by a, b, c, d, respectively, the relative frequencies p, p' in the two rows as observed could have no frequency distributions because $N + 1$ of the $(N+3)(N+2)(N+1)/6$ samples would turn out with $a + b = 0$ and $N + 1$ of them with $c + d = 0$ and the ratios for p or p' would be $0/0$; the total probability that $p - p'$ would be indeterminate would ordinarily be small but the fact that there would be indetermination is sufficient to show that there would logically be no distribution for $p - p'$ and hence no moments.[4] However, the expression $x = (ad - bc)/N$, though no longer interpretable in terms of the difference $p - p'$ would have a distribution, not of $c + d + 1$ elements nor of $(a + b + 1)(c + d + 1)$ but of $(N+3)(N+2)(N+1)/6$. The value of μ_2 for this distribution is readily found to be[5]

$$\mu_2(x) = (N-1)p_1 p_2 q_1 q_2 \quad \text{compared with } N p_1 p_2 q_1 q_1 \quad \text{or } \frac{N^2}{N-1} p_1 p_2 q_1 q_2$$

when the sampling is by the double point binomial or by the Yates-Fisher series, respectively, the values of the unknown probabilities in the universes in question being taken from the marginal totals.

The argument by which the linear series of tables with their relative probabilities are obtained from a given 2×2 table by considering variations in the body of the table consistent with the given marginal totals is equally applicable to $m \times n$ tables.[6] Given the table

$$\begin{array}{c|c} a_{11}\ a_{12}\ \dots\ a_{1m} & A_1 \\ a_{21}\ a_{22}\ \dots\ a_{2m} & A_2 \\ \dots\dots\dots\dots & \\ a_{n1}\ a_{n2}\ \dots\ a_{nm} & A_n \\ \hline B_1\ B_2 \quad B_m & N \end{array} \quad \text{then} \quad \frac{\Pi_i A_i!\Pi_j B_j!}{N!}\left[\frac{1}{\Pi_{ij} a_{ij}!}\right]$$

if Π be the sign of a product, is its relative probability in the $(m-1)(n-1)$ dimensional spread[7] of tables derived therefrom by all variations of a_{ij} consistent with the marginal totals A_i and B_j. The mean value in such a spread for the term a_{ij} is A_iB_j/N. The average value of $a_{ij}(a_{ij}-1)$ or of $a_{ij}a_{ik}$, $k \neq j$, or of $a_{ij}a_{kl}$, $k \neq i$, $l \neq j$, can be found by an argument entirely similar to that given before. For example,

$$\overline{a_{12}a_{34}} = \frac{\Pi_i A_i!\Pi_j B_j!}{N!}\left[\sum \frac{a_{12}a_{34}}{\Pi_{ij} a_{ij}!} = \sum{}' \frac{1}{(a_{12}-1)!(a_{34}-1)!\Pi' a_{ij}!}\right]$$

where $\sum'$ extends to all variations consistent with the former marginal totals except that A_1 is replaced by $A_1 - 1$, B_2 by $B_2 - 1$, A_3 by $A_3 - 1$, B_4 by $B_4 - 1$ and Π' extends over the $mn - 2$ elements other than a_{12} and a_{34}.

$$\overline{a_{12}a_{34}} = \frac{A_1B_2A_3B_4}{N(N-1)}$$

$$\overline{a_{12}a_{34}} - \bar{a}_{12}\bar{a}_{34} = \frac{A_1B_2A_3B_4}{N(N-1)} - \frac{A_1B_2}{N}\,\frac{A_3B_4}{N}$$

or the mean product relative to the mean is

$$\pi(a_{ij}a_{kl}) = \frac{N^2}{N-1}\, p_iq_j\, p_kq_l,\ k \neq i,\ l \neq j$$

where $p_i = A_i/N$, $q_j = B_j/N$. Similarly

$$\mu_2(a_{ij}) = \frac{N^2}{N-1}\, p_i(1-p_i)q_j(1-q_j)$$

$$\pi(a_{ij}a_{il}) = -\frac{N^2}{N-1}\, p_i(1-p_i)q_jq_l.$$

Hence the correlation coefficients of two terms not in the same row or column and of two terms in the same row but not in the same column are

$$r(a_{ij}a_{kl}) = +\sqrt{\frac{p_iq_jp_kq_l}{(1-p_i)(1-q_j)(1-p_k)(1-q_l)}},$$

$$r(a_{ij}a_{il}) = -\sqrt{\frac{q_jq_l}{(1-q_j)(1-q_l)}}.$$

Higher moments could be obtained, with sufficient patience, if they were needed.

It sometimes happens that the numbers in two frequency functions are really small. For example, the distribution by age of deaths from chicken pox for the two sexes as reported in Massachusetts for 1934 were

Age	<1	1	2	3	4	5–9	10–14	15–19	20–29	30–39	40–49	50–59	Total
Male	3	1	0	0	0	1	0	0	0	1	0	1	7
Female	0	2	1	0	2	0	0	0	0	0	0	0	5
Total	3	3	1	0	2	1	0	0	0	1	0	1	12

with none for all higher age groups. This as it stands without the higher age groups is a 2×12 table; but it seems rather clear that it has to be considered as a 2×7 table, omitting the columns with no elements.[8] If so considered $\chi^2 = 9.28$ looked up under 6 degrees of freedom gives $P = 0.17$ which is not significant. There are 130 different tables consistent with the marginal totals. The relative probability of the given table is 1/264. The 130 relative probabilities which arise have the frequency distribution

prob.	$\frac{1}{792}$	$\frac{2}{792}$	$\frac{3}{792}$	$\frac{6}{792}$	$\frac{9}{792}$	$\frac{18}{792}$
freq.	18	9	40	22	26	15

The total probability of the given table and all no more probable is $52/264 = 0.20$ which is not much larger than the value previously obtained from χ^2. However, the variance of the age for the males is 478 with 6 degrees of freedom and for the females is 2.4 with 4 degrees of freedom which gives a ratio of 208 and the 0.01 point is at 15.2. It is far from certain that variance analysis can be relied upon; but almost any test for the significance of the standard errors of the distributions would indicate a high degree thereof—thus indicating how different the result of different tests corresponding to different formulations of the problem may be.

[1] Yates, F., "Contingency Tables Involving Small Numbers and the χ^2 Test." *Suppl. Jour. Roy. Statist. Soc.*, 1, 217–235 (1934).

[2] The tables $\frac{n-k}{k}\Big|\frac{k}{n-k}$ and $\frac{n-k}{k}\Big|\frac{k}{n-k-1}$ where $k < {}^1/_2 n$ have the interesting property that the total probability of the given tables and of all tables in the respective series which are no more probable than these (taking both tails together) are equal and equal to

$$\frac{2(n!)^2}{(2n)!}\left\{1^2 + n^2 + \left[\frac{n(n-1)}{2!}\right]^2 + \ldots + \left[\frac{n(n-1)\ldots(n-k+1)}{k!}\right]^2\right\}$$

Thus $\frac{4}{0}\Big|\frac{0}{4}$ and $\frac{0}{4}\Big|\frac{4}{0}$ together have the same probability $\frac{1}{35}$ as $\frac{4}{0}\Big|\frac{0}{3}$ and the tables $\frac{5}{1}\Big|\frac{1}{5}, \frac{6}{0}\Big|\frac{0}{6}$, $\frac{1}{5}\Big|\frac{5}{1}, \frac{0}{6}\Big|\frac{6}{0}$ have the same probability as $\frac{5}{1}\Big|\frac{1}{4}, \frac{6}{0}\Big|\frac{0}{5}, \frac{1}{5}\Big|\frac{5}{0}$, namely, 37/462.

[3] As an illustrative example consider the table $a = 10, b = 3, c = 2, d = 15, N = 30$. The 13 relative probabilities are as 1; 102; 2992; 37,400; 235,620; 816,816; 1,633,632; 1,925,352; 1,337,050; 534,820; 116,688; 12,376; 476, the common denominator being 6,653,325. The series looks skew in the arithmetical sense, but actually $\kappa = 0.02113$; also $\beta_2 - 3 = -0.08084$. If we note that for a point binomial the skewness κ is $(q - p)/\sqrt{Npq}$ we note that for x in the Yates-Fisher series $\kappa = \sqrt{N-1}(1 - 2/N)\kappa_1\kappa_2$ and thus varies for large values of N jointly as the square root of N and the product of the skewnesses of the point binomials in the margins.

[4] If $N = 100$, and $p_1 = 0.02$ for the universe, p_2 being anything, there would be 2 elements on the average in the first row which would, however, be entirely vacant in e^{-2} or 13.5% of the samples.

[5] Had we assumed a general four-fold universe the average value of $x = (ad - bc)/N$ and of its variance would have been

$$\bar{x} = (N-1)(\pi_1 - p_1p_2), \qquad \mu_2(x) = (N-1)\left[p_1p_2q_1q_2 - \left(1 - \frac{2}{N}\right)\delta^2 + \left(1 - \frac{1}{N}\right)\delta(q_1 - p_1)(q_2 - p_2)\right]$$

where

$$\delta = \pi_1 - p_1p_2 \text{ and the universe is } \frac{\pi_1}{p_2 - \pi_1}\left|\frac{p_1 - \pi_1}{1 - p_1 - p_2 + \pi_1}\right.$$

[6] Let it be assumed that $q_1, q_2, \ldots, q_m$ are the probabilities that an element fall in column 1, 2, ..., m, respectively, in any row. Then the probability that $a_{11}, a_{12}, \ldots, a_{1m}$, respectively, will fall in the n cells of the first row is

$$\frac{A_1!q_1^{a_{11}}q_2^{a_{12}} \ldots q_m^{a_{1m}}}{a_{11}!a_{12}! \ldots a_{1m}!} \text{ and } \frac{\Pi_iA_i!}{\Pi_ja_{.j}!}q_1^{B_1}q_2^{B_2} \ldots q_m^{B_m}$$

will be the probability of the $m \times n$ observed frequencies in the table. The unknown factor involving q_i will be the same for all tables having the same marginal totals. But the probability of those marginal totals arising from the distribution of N elements is

$$\frac{N!}{\Pi_jB_j!}q_1^{B_1}q_2^{B_2} \ldots q_m^{B_m} \text{ and hence the quotient } \frac{\Pi_iA_i!\Pi_jB_j!}{N!\Pi_{ij}a_{ij}!}$$

is the relative probability for any distribution a_{ij} consistent with the marginal totals.

[7] Whether "dimensions" or "degrees of freedom" or "spread" are proper terms to use for variations when numbers are very small may be questioned. One may illustrate the procedure for a particular case. Consider

$$\frac{3, 0, 2}{0, 3, 0} \text{ and its relative probability } \frac{5!(3!)^32!}{8!}\frac{1}{(3!)^32!(0!)^3} = \frac{1}{56}$$

together with all its variants and their probabilities (in brackets)

$$\frac{3, 1, 1}{0, 2, 1}\left[\frac{6}{56}\right]; \frac{3, 2, 0}{0, 1, 2}\left[\frac{3}{56}\right]; \frac{2, 1, 2}{1, 2, 0}\left[\frac{9}{56}\right]; \frac{2, 2, 1}{1, 1, 1}\left[\frac{18}{56}\right];$$

$$\frac{2, 3, 0}{1, 0, 2}\left[\frac{3}{56}\right]; \frac{1, 3, 1}{2, 0, 1}\left[\frac{6}{56}\right]; \frac{1, 2, 2}{2, 1, 0}\left[\frac{9}{56}\right]; \frac{0, 3, 2}{3, 0, 0}\left[\frac{1}{56}\right].$$

The probability of the given table and all no more probable ones is 2/56, and if one uses a level of 0.05 for significance, and uses the criterion of the given table and all no more probable ones, this table would be significant. Nothing has been said of the sampling

process by which the table arose and if I understand Yates[1] aright it makes no difference whether it arose by distributing 8 elements over the six cells from a universe with the probabilities $\frac{\pi_1}{\pi_4}\left|\frac{\pi_2}{\pi_5}\right|\frac{\pi_3}{\pi_6}$ with $\Sigma\pi = 1$ or by distributing 5 elements and 3 elements, respectively, to each of two trinomial universes for which $\pi_1 + \pi_2 + \pi_3 = \pi_4 + \pi_5 + \pi_6 = 1$ or by distributing 3, 3 and 2 elements, respectively, to each of three binomial universes for which $\pi_1 + \pi_4 = \pi_2 + \pi_5 = \pi_3 + \pi_6 = 1$.

[8] If the tabulation were by sufficiently fine age groups the two rows of the contingency table would have nothing but 0's or 1's and so would the totals in the columns. If there were 7 entries of 1 in the top row and 5 entries of 1 in the bottom row and 12 entries of 1 in the marginal totals along the bottom, there would be $12!/(7!5!) = 792$ different tables all of equal probability 1/792, and the total probability of the no more probable tables would be 1.

CONCERNING THE DIFFERENTIAL EQUATIONS OF SOME BOUNDARY LAYER PROBLEMS. II

BY HERMANN WEYL

INSTITUTE FOR ADVANCED STUDY

Communicated January 29, 1942

In a previous note under the same title[1] I have solved by successive approximations the following boundary value problem:

$$(A_\lambda)\quad \begin{cases} w''' + 2ww'' + 2\lambda(k^2 - w'^2) = 0 \text{ for } z \geqq 0; \\ w = w' = 0 \text{ for } z = 0,\ w'(\infty) = k, \end{cases}$$

involving the positive parameter k, for the special values $\lambda = 0$ and $1/2$. Convergence of the process for $\lambda = 1/2$ requires a certain constant q to be < 1. The actual computation of q is laborious; however, I wish to report that a comparatively simple estimation yields $q < 3/4$.

For *arbitrary* $\lambda \geqq 0$ I have to take refuge to a much more sophisticated method, that of fixed points of functional transformations, which was inaugurated by Birkhoff and Kellogg in a famous memoir[2] and later formulated in general abstract terms by Schauder and Leray. The experiences with the special cases suggest the following procedure. We eliminate the parameter k from the differential equation by differentiation,

$$w'''' + 2ww''' + 2(1 - 2\lambda)w'w'' = 0,$$

and observe that the function $\kappa.w(\kappa z)$ solves the new equation of fourth order if $w(z)$ does, whatever the positive constant κ. Hence we seek our solution of (A_λ) in the form $w(z) = \kappa.f(\kappa z)$ where

$$\left.\begin{aligned} f'''' + 2ff''' + 2(1 - 2\lambda)f'f'' = 0 \text{ for } z \geqq 0; \\ f(0) = f'(0) = 0, f''(0) = 1;\ f''(\infty) = 0. \end{aligned}\right\} \qquad (1)$$

Our line of attack can be explained a little more easily if we replace the infinite by the finite interval $0 \leqq z \leqq a$ and thus the last boundary condition in (1) by $f''(a) = 0$. We then introduce $f'' = g = \varphi$ as the unknown function and start with this set-up:

$$f(z) = \int_0^z (z - \zeta) g(\zeta) d\zeta. \tag{2}$$

$$\left.\begin{aligned} \varphi'' + 2f\varphi' + 2(1 - 2\lambda) f'\varphi = 0; \\ \varphi(0) = 1, \varphi(a) = 0. \end{aligned}\right\} \tag{3}$$

$$\varphi = g. \tag{4}$$

For an arbitrarily given g we form (2), then solve the linear boundary value problem (3) of familiar type, thus defining the functional operator Φ_λ which carries g into φ, and as the last step (4), we try to determine g as a function whose image coincides with the original,

$$g = \Phi_\lambda\{g\}.$$

The success hinges on the following

LEMMA: *Under the assumption* $g(z) \geqq 0$ *the linear problem* (3) *has a unique solution* φ; *it satisfies the conditions*

$$\varphi \geqq 0, \varphi' + 2f\varphi \leqq 0 \tag{5}$$

throughout the interval $0 \leqq z \leqq a$.

To prove this we set

$$p(z) = \exp(2\int_0^z f(\zeta) d\zeta)$$

so that $p(0) = 1$, $p' = 2fp$, and introduce the auxiliary function

$$\varphi_1 = p(\varphi' + 2f\varphi) = (p\varphi)'.$$

The single differential equation for φ is then replaced by the system

$$\left.\begin{aligned} \varphi' + 2f\varphi = \varphi_1/p, \\ \varphi_1' - 2f\varphi_1 = 4\lambda p f' \varphi. \end{aligned}\right\} \tag{6}$$

Multiplication by φ_1 and φ, respectively, and addition leads to the formula

$$[\varphi\varphi_1]_b^a = \int_b^a \left(\frac{1}{p}\varphi_1^2 + 4\lambda p f' \varphi^2\right) dz \qquad (b < a). \tag{7}$$

The inequality $g \geqq 0$ which implies $f' \geqq 0$ guarantees the positive definite character of the "Dirichlet integral" at the right side of (7). Our lemma is an easy consequence of this fact.

Thus we have defined the operator Φ_λ for all $g \geqq 0$; the restriction $g \geqq 0$ is harmless because it carries over to the image φ. Since (5) implies $(p\varphi)' \leqq 0$ we have even

$$p\varphi \leqq 1, 0 \leqq \varphi \leqq 1,$$

and may therefore start with the set G of all functions $g(z)$ defined and continuous in the interval $0 \leqq z \leqq a$ for which $0 \leqq g \leqq 1$. This limitation of g has the consequence that

$$0 \leqq f' \leqq z, 0 \leqq f \leqq 1/2z^2 \tag{8}$$

and enables us to ascertain a universal bound for φ', i.e., a number A depending on λ and a only such that

$$0 \leqq -\varphi'(z) \leqq A$$

for $0 \leqq z \leqq a$ and the image $\varphi = \Phi_\lambda\{g\}$ of *every* $g \epsilon G$. First we determine such a universal bound for $\varphi'(0)$. Indeed would φ start off at $z = 0$ with too steep a decline it could not help hitting the ground, $\varphi = 0$, long before z reaches a, considering that the coefficients of its differential equation (3) are bounded by (8). Once $\varphi'(0)$ has been penned in that differential equation does the rest. Hence the images φ of all elements $g \epsilon G$ are equi-continuous in the strict sense that $|z_1 - z_2| \leqq \epsilon$ implies $|\varphi(z_1) - \varphi(z_2)| \leqq A\epsilon$.

We topologize the "space" of all continuous functions of z by interpreting convergence as *uniform* convergence in the interval $0 \leqq z \leqq a$. Then Φ_λ is a continuous operator in G which maps G into the subset G_A consisting of all functions $g \epsilon G$ for which $|z_1 - z_2| \leqq \epsilon$ implies $|g(z_1) - g(z_2)| \leqq A\epsilon$. This suffices for the existence of a "fixed point" $g \epsilon G$ of the operator Φ_λ (see Birkhoff and Kellogg, l.c., Theorem II).

The case of the infinite interval is not essentially more difficult. It turns out that the solution f of (1) which our method yields has the property that $f'' \geqq 0$ and $f''' + 2ff'' \leqq 0$ converge to zero with $z \to \infty$ at least as strongly as a function of the type $e^{-\gamma z^2}(\gamma > 0)$. Hence

$$\int_0^\infty f''(z)dz = f'(\infty) = \beta$$

converges and we may retrace the step leading from (A_λ) to (1).

THEOREM. *For given positive k the problem (A_λ) has a solution w whose derivative is monotone increasing from 0 to k as z travels from 0 to infinity; the second derivative decreases monotonely from*

$$w''(0) = \alpha k^{3/4} \quad (\alpha = \beta^{-3/4})$$

to zero, approaching zero with $z \to \infty$ at least as strongly as a function of the type $e^{-\gamma z^2}$.

[1] These PROCEEDINGS, **27**, 578–583 (1941).

[2] Birkhoff, G. D., and Kellogg, O. P., *Trans. Am. Math. Soc.*, **23**, 96–115 (1922); Schauder, J., *Studia Mathematica*, **2**, 171–180 (1930); Leray, J., and Schauder, J., *Ann. Sc. Ec. Norm. Sup.*, **51**, 45–78 (1934).

UNSTABLE MINIMAL SURFACES WITH ANY RECTIFIABLE BOUNDARY

BY MAX SHIFFMAN

COLLEGE OF THE CITY OF NEW YORK

Communicated December 17, 1941

1. Introduction.—Let Γ be any rectifiable curve in space. We shall prove that Γ must bound unstable minimal surfaces whenever it bounds several minimal surfaces which are relative minima in area. More generally, the Morse relations apply in a slightly modified form. In previous literature, these results were obtained for special types of rectifiable curves.[1]

The method is first to obtain unstable minimal surfaces bounded by curves approximating Γ and second to perform a passage to the limit. The first step may be done in several ways. In the present paper it is carried through by extending the procedure used by Courant for polygonal boundaries.[2] The passage to the limit is accomplished by means of theorem 1 below, a theorem which was discovered by Morse and Tompkins,[3] and which is here proved by using the isoperimetric inequality for minimal surfaces.[4]

Our method has the advantage of not requiring any explicit representation of the Dirichlet integral of a potential surface in terms of its boundary values.

2. Continuity of Area of Minimal Surfaces.—2.1. Let $\xi(r, \theta)$ be a potential surface defined over the unit circle $r \leq 1$, with boundary values $\xi(1, \theta)$ which are continuous and of bounded variation. Indicate the length of the curve $\xi(r, \theta)$, for fixed r and $0 \leq \theta \leq 2\pi$, by $L(r)$; indicate the length of $\xi(r, \theta)$ for fixed θ and $\rho \leq r \leq 1$ by $L(\theta;\ \rho, 1)$. The symbol $D[\xi]$ stands for the Dirichlet integral

$$D[\xi] = \frac{1}{2}\int\int\left(\xi_r{}^2 + \frac{1}{r^2}\xi_\theta{}^2\right)rdrd\theta.$$

LEMMA 1: $L(r) \leq L(1)$.

LEMMA 2: *Let $\xi^n(r, \theta)$, $n = 1, 2, \ldots$, be a sequence of potential surfaces converging uniformly[5] to the potential surface $\xi^\infty(r, \theta)$, and suppose that $L^n(1) \to L^\infty(1)$ as $n \to \infty$. Then for any ϵ there is a δ such that any arc concentric to the unit circle whose angle is less than δ is mapped by each ξ^n, $n = 1, 2, \ldots, \infty$, into curves of length less than ϵ.*

LEMMA 3: $L(\theta;\ \rho,\ 1) \leq \left(\frac{4D[\xi]}{\delta} \cdot \log\frac{1}{\rho}\right)^{1/2}$ *except for a set of θ's of total measure $< \delta/2$.*

Lemma 1 is proved by use of the Poisson integral. Lemma 2 is a conse-

quence of Lemma 1 and the lower semi-continuity of length. Lemma 3 is obtained by using well-traversed estimates of $D[\mathfrak{x}]$.[6]

Define a *δ-dense radial network* in the annular ring $\rho \leq r \leq 1$ as a set of at most $\left[\frac{4\pi}{\delta}\right] + 1$ radial lines $\theta =$ constant, such that two successive radial lines have an angular difference of at most δ. Lemma 3 shows that there is a δ-dense radial network in the annular ring $\rho \leq r \leq 1$ on each line of which the inequality of Lemma 3 holds.

2.2. We consider a minimal surface $\mathfrak{x}(r, \theta)$ as given in isometric representation, i.e., $\mathfrak{x}$ is a potential surface and $E = G$, $F = 0$ where $E = \mathfrak{x}_r^2$, $F = \mathfrak{x}_r \cdot 1/r\, \mathfrak{x}_\theta$, $G = 1/r^2\, \mathfrak{x}_\theta^2$. The area of a minimal surface is equal to its Dirichlet integral.

THEOREM 1.[7] *Let $\mathfrak{x}^n(r, \theta)$, $n = 1, 2, \ldots$, be a sequence of minimal surfaces converging uniformly[5] to the minimal surface $\mathfrak{x}^\infty(r, \theta)$, and suppose that $L^n(1) \to L^\infty(1)$ as $n \to \infty$. Then*

$$D[\mathfrak{x}^n] \to D[\mathfrak{x}^\infty] \text{ as } n \to \infty.$$

Proof. Let M be an upper bound of the quantities $L^n(1)$, $[L^n(1)]^2/4\pi$, $n = 1, 2, \ldots, \infty$. By the isoperimetric inequality $A \leq L^2/4\pi$ for minimal surfaces, we have $D[\mathfrak{x}^n] \leq M$ for all n.

For any given ϵ choose the δ of Lemma 2, and then ρ so near 1 that

$$\frac{2M}{\pi}\left(\frac{4M}{\delta} \cdot \log\frac{1}{\rho}\right)^{1/2} + \left(\frac{4\pi}{\delta} + 1\right)\frac{4M}{\pi\delta} \cdot \log\frac{1}{\rho} < \epsilon.$$

Consider any $\mathfrak{x}^n$, $n = 1, 2, \ldots, \infty$. For this $\mathfrak{x}^n$, there is a δ-dense radial network in the annular ring $\rho \leq r \leq 1$ satisfying the inequality of Lemma 3. Consider the ith cell of this network. Indicate the Dirichlet integral of $\mathfrak{x}^n$ over this cell by D_i'; and indicate the lengths of the images under $\mathfrak{x}^n$ of the four boundary lines of this cell by a_i, c_i, b_i, c_{i-1}, where c_i, c_{i-1} correspond to the radial lines, a_i to the arc of the unit circle and b_i to the arc of the circle of radius ρ. Set $l_i = a_i + b_i + c_i + c_{i-1}$. We have a_i, $b_i < \epsilon$ by Lemma 2 and the choice of δ, and c_i, $c_{i-1} < \left(\frac{4M}{\delta} \cdot \log\frac{1}{\rho}\right)^{1/2}$. By the isoperimetric inequality,

$$4\pi D_i' \leq l_i^2 = (l_i + c_i + c_{i-1})(a_i + b_i) + (c_i + c_{i-1})^2 \leq$$

$$\leqslant \left[2\epsilon + 4\left(\frac{4M}{\delta} \cdot \log\frac{1}{\rho}\right)^{1/2}\right](a_i + b_i) + \frac{16M}{\delta}\log\frac{1}{\rho} \quad (1)$$

Denote the Dirichlet integral of $\mathfrak{x}^n$ over the annular ring $\rho \leq r \leq 1$ by $D'[\mathfrak{x}^n]$; over the circle $r \leq \rho$ by $D''[\mathfrak{x}^n]$. Summing the inequality (1) for at most $[4\pi/\delta] + 1$ terms, and making use of $\sum a_i = L^n(1) \leq M$ and $\sum b_i = L^n(\rho) \leq L^n(1) \leq M$, we obtain

$$D'[\mathfrak{x}^n] \leq \frac{M\epsilon}{\pi} + \frac{2M}{\pi}\left(\frac{4M}{\delta}\cdot\log\frac{1}{\rho}\right)^{1/2} + \left(\frac{4\pi}{\delta}+1\right)\frac{4M}{\pi\delta}\log\frac{1}{\rho} < (M+1)\epsilon. \quad (2)$$

This inequality holds for $n = 1, 2, \ldots, \infty$.

Choose N so large that $|D''[\mathfrak{x}^n] - D''[\mathfrak{x}^\infty]| < \epsilon$ for all $n \geq N$. Inequality (2) yields

$$|D[\mathfrak{x}^n] - D[\mathfrak{x}^\infty]| < (2M+3)\epsilon \text{ for all } n \geq N.$$

The proof of the theorem is completed.

3. Unstable Minimal Surfaces Bounded by Curves Approximating Γ.—3.1. Let $\mathfrak{P}$ be the space of potential surfaces $\mathfrak{x}(r, \theta)$ defined over the unit circle, having a finite Dirichlet integral, mapping the circumference monotonically and continuously onto Γ, and satisfying the three point condition. We define the distance between two surfaces $\mathfrak{x}_1$, $\mathfrak{x}_2$ as[8]

$$|\mathfrak{x}_1 - \mathfrak{x}_2| = \underset{r \leq 1/2}{\text{Max.}} |\mathfrak{x}_1(r, \theta) - \mathfrak{x}_2(r, \theta)|.$$

We shall first obtain unstable minimal surfaces in a space $\bar{\mathfrak{R}}$ containing $\mathfrak{P}$, and then take the limit as $\bar{\mathfrak{R}} \to \mathfrak{P}$.

Let $A = \{A_1, A_2, \ldots, A_m\}$ be a collection of m distinct points arranged in this order on the oriented curve Γ. Designate the maximum among the diameters of the arcs $A_1A_2, A_2A_3, \ldots, A_mA_1$ of Γ by $\sigma(A)$. The collection A is fixed in this section.

Let $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ be a collection of m distinct points arranged in this order on the circumference. Define the following three sets of potential surfaces:

(i) $\mathfrak{P}(\alpha)$ is the set of all surfaces $\mathfrak{x}$ of $\mathfrak{P}$ for which $\mathfrak{x}(\alpha_j) = A_j, j = 1,2,\ldots, m$.

(ii) $\mathfrak{R}(\alpha)$ is the set of all convex linear combinations $\mathfrak{y}(r, \theta)$ of surfaces in $\mathfrak{P}(\alpha)$, i.e., $\mathfrak{y}$ is expressible as

$$\mathfrak{y} = \sum a_i\mathfrak{x}_i \text{ where } a_i \geq 0, \sum a_i = 1, \text{ and each } \mathfrak{x}_i \text{ is in } \mathfrak{P}(\alpha).$$

(iii) $\bar{\mathfrak{R}}(\alpha)$ is the set of all surfaces $\mathfrak{z}(r, \theta)$ with finite Dirichlet integral obtainable as limits of sequences of surfaces in $\mathfrak{R}(\alpha)$, i.e., $D[\mathfrak{z}] < \infty$ and $\mathfrak{z} = \lim \mathfrak{y}^n$ where each $\mathfrak{y}^n$ is in $\mathfrak{R}(\alpha)$. The set $\bar{\mathfrak{R}}(\alpha)$ may be described as the *convex closure* of the set $\mathfrak{P}(\alpha)$.

Concerning $\bar{\mathfrak{R}}(\alpha)$, the following statements are easily proved.

(*a*) $\mathfrak{P}(\alpha) \subset \mathfrak{R}(\alpha) \subset \bar{\mathfrak{R}}(\alpha)$.

(*b*) If $\mathfrak{z}$ is in $\bar{\mathfrak{R}}(\alpha)$, $\mathfrak{z}(1, \theta)$ is a continuous curve whose length is at most the length L of Γ; and $\mathfrak{z}(1, \theta)$ for $\alpha_{i-1} \leq \theta \leq \alpha_i$ lies in the convex closure

of the arc $\overset{\frown}{A_{i-1}A_i}$ of Γ. If χ is any surface in $\mathfrak{P}(\alpha)$, $\max_{r \leq 1} |\zeta(r, \theta) - \chi(r, \theta)| \leq \sigma(A)$.

(*c*) The set $\overline{\mathfrak{R}}(\alpha)$ is convex, i.e., if ζ_1, ζ_2 belong to $\overline{\mathfrak{R}}(\alpha)$ so does $\zeta = (1 - t)\zeta_2 + t\zeta_1$ for $0 \leq t \leq 1$.

(*d*) The set $\overline{\mathfrak{R}}(\alpha)$ is D-compact, i.e., if ζ^n is a sequence belonging to $\overline{\mathfrak{R}}(\alpha)$ for which $D[\zeta^n] \leq M$, a subsequence can be found which converges to a surface ζ in $\overline{\mathfrak{R}}(\alpha)$.

(*e*) Let ζ belong to $\overline{\mathfrak{R}}(\alpha)$, and vary ζ by setting

$$Z'(r, \Theta) = \zeta(r, \theta) \text{ for } \Theta = \theta + \epsilon\lambda(r, \theta)$$

where $\lambda(r, \theta)$ has continuous first derivatives and ϵ is chosen so small that $|\epsilon\lambda_\theta| \leq 1/2$. Transform Z' by a linear transformation so that the three point condition is satisfied, and define ζ' as the potential surface having the same boundary values as the resulting surface. Set $\alpha' =$ image of α under these successive transformations. Then ζ' belongs to $\overline{\mathfrak{R}}(\alpha')$.

3.2. We are ready to define the space $\overline{\mathfrak{R}}$. $\overline{\mathfrak{R}}$ consists of all the surfaces in the $\overline{\mathfrak{R}}(\alpha)$ for all possible positions of the collection $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$. The distance between two surfaces ζ_1, ζ_2 of $\overline{\mathfrak{R}}$ is again defined by

$$|\zeta_1 - \zeta_2| = \max_{r \leq 1/2} |\zeta_1(r, \theta) - \zeta_2(r, \theta)|.$$

By property 3.1 (*a*), $\mathfrak{P} \subset \overline{\mathfrak{R}}$. Property 3.1 (*e*) shows that it is possible to perform variations in the space $\overline{\mathfrak{R}}$.

By using properties 3.1 (*c*), (*d*), (*e*) and methods similar to [3] the following theorems can be proved:

THEOREM 2. *The problem of minimizing $D[\zeta]$ among all surfaces ζ in $\overline{\mathfrak{R}}(\alpha)$ has a unique solution, denoted by $\zeta(\alpha)$.*

THEOREM 3. *The surface $\zeta(\alpha)$ and the quantity $d(\alpha) = D[\zeta(\alpha)]$ depend continuously on the collection $\alpha = \{\alpha_1, \ldots, \alpha_m\}$.*

As in [3] it may also be proved that $d(\alpha)$ is a differentiable function of the $\alpha_1, \alpha_2, \ldots, \alpha_m$.

Define the subspace Ω of $\overline{\mathfrak{R}}$ as the set of surfaces $\zeta(\alpha)$ for all possible α.

Theorem 2 and property 3.1 (*c*) assert that $\overline{\mathfrak{R}}$ may be D-retracted into Ω, i.e., retracted in such a way that the Dirichlet integral never increases. Theorem 3 asserts that Ω is topologically equivalent to the space of all possible α's, and that the Dirichlet integral is continuous in Ω.

Thus, so far as Morse theory in $\overline{\mathfrak{R}}$ is concerned, we may limit ourselves to the space Ω; and Morse theory applies to the minimal surfaces in Ω.

In particular, let χ_1, χ_2 be any two surfaces in $\overline{\mathfrak{R}}$. The surfaces χ_1, χ_2 can both be D-retracted into surfaces ζ_1, ζ_2 in Ω. In Ω, ζ_1 and ζ_2 can be joined by a minimizing connected set C; the maximum d of the Dirichlet integral on C is the smallest possible among all connected sets in Ω joining ζ_1, ζ_2. If $d > \max.\{D[\chi_1], D[\chi_2]\}$, then C contains a minimal surface ζ for which

$D[\mathfrak{z}] = d$. Because of the isoperimetric inequality for minimal surfaces, we have $d \leq L^2/4\pi$ by 3.1 (*b*).

THEOREM 4. *Two surfaces* $\mathfrak{x}_1, \mathfrak{x}_2$ *in* $\mathfrak{R}$ *can be joined by a connected set C of surfaces in* $\overline{\mathfrak{R}}$ *on which the maximum of* $D[\mathfrak{z}]$ *is a quantity* $d \leq max.\{D[\mathfrak{x}_1], D[\mathfrak{x}_2], L^2/4\pi\}$. *If* $d > max.\{D[\mathfrak{x}_1], D[\mathfrak{x}_2]\}$, *then C contains a minimal surface* $\mathfrak{z}$ *for which* $D[\mathfrak{z}] = d$.

4. *Passage to the Limit to Yield Unstable Minimal Surfaces Bounded by* Γ. —4.1. The considerations in §3 depended on the selection of the collection of points $A = \{A_1, A_2, \ldots, A_m\}$ on Γ. The space $\overline{\mathfrak{R}}$ defined there will now be denoted by $\overline{\mathfrak{R}}(A)$.

Select a sequence $A^{(1)}, A^{(2)}, \ldots, A^{(\nu)}, \ldots$ of collections of points on Γ, where $A^{(\nu+1)}$ is obtained from $A^{(\nu)}$ by adding points of Γ, and where $\sigma(A^{(\nu)}) \to 0$ as $\nu \to \infty$. Then $\overline{\mathfrak{R}}(A^{(1)}) \supset \overline{\mathfrak{R}}(A^{(2)}) \supset \ldots \supset \overline{\mathfrak{R}}(A^{(\nu)}) \supset \ldots \supset \mathfrak{P}$. And using 3.1 (*b*), $\lim_{\nu \to \infty} \overline{\mathfrak{R}}(A^{(\nu)}) = \mathfrak{P}$.

4.2. MAIN THEOREM I. *If the rectifiable closed Jordan curve* Γ *bounds two minimal surfaces which are proper relative minima, then* Γ *must bound an unstable minimal surface.*

Proof. Let the two minimal surfaces be $\mathfrak{x}_1, \mathfrak{x}_2$. By Theorem 4, there is a connected set $C^{(\nu)}$ of surfaces in $\overline{\mathfrak{R}}(A^{(\nu)})$ joining $\mathfrak{x}_1, \mathfrak{x}_2$ on which the maximum of $D[\mathfrak{z}]$ is a quantity $d^{(\nu)}$ satisfying the inequality of Theorem 4. Select a sequence of $A^{(\nu)}$ for which the $d^{(\nu)}$ converge, and indicate it still by $A^{(\nu)}$. Consider all the limiting surfaces of the sets $C^{(\nu)}$ as $\nu \to \infty$. They form a connected set C in $\mathfrak{P}$ containing $\mathfrak{x}_1, \mathfrak{x}_2$ on which the maximum of $D[\mathfrak{x}]$ is $d \leq \lim_{\nu \to \infty} d^{(\nu)} \leq L^2/4\pi$.

Now, $d > \max.\{D[\mathfrak{x}_1], D[\mathfrak{x}_2]\}$ because $\mathfrak{x}_1, \mathfrak{x}_2$ are *proper* relative minima. Hence $d^{(\nu)} > \max.\{D[\mathfrak{x}_1], D[\mathfrak{x}_2]\}$ for all ν sufficiently large. By Theorem 4 there is a minimal surface $\mathfrak{z}^\nu$ on $C^{(\nu)}$ for which $D[\mathfrak{z}^\nu] = d^{(\nu)}$. A subsequence of the $\mathfrak{z}^\nu$ converges to a minimal surface $\mathfrak{z}$ on C. The lower semi-continuity of length and property 3.1 (*b*) show that the lengths of the curves $\mathfrak{z}^\nu(1, \theta)$ converges to the length L of Γ. Theorem 1 yields

$$\lim_{\nu \to \infty} d^\nu = D[\mathfrak{z}] \leq d.$$

Hence $d = \lim_{\nu \to \infty} d^{(\nu)}$ and $D[\mathfrak{z}] = d$. The minimal surface required by Main Theorem I is $\mathfrak{z}$.

5. *The Morse Relations.*—In the same way, by enlarging the space to $\overline{\mathfrak{R}}(A^{(\nu)})$ and then passing to the limit $\mathfrak{P}$, one can prove the following:

(1) any k-cap in $\mathfrak{P}$ with cap limit a is homologous on $\mathfrak{P}_a$ to a k-cap which contains a minimal surface $\mathfrak{z}$ for which $D[\mathfrak{z}] = a$;[9]

(2) the connectivity numbers of $\mathfrak{P}$ are $R_0 = 1$, $R_1 = R_2 = \ldots = R_a = \ldots = 0$.

In these two statements, we always consider Vietoris cycles lying on $\mathfrak{P}_N$

for some N. The only difficulty in the proof of these statements, beyond the methods already used in §§1–4, is the possibility that surfaces in $\mathfrak{P}$ may have boundary values which are constant on some arcs of the circumference.

Statement (1) yields a modified Morse theory; the usual requirement is that every k-cap contain a minimal surface $\mathfrak{z}$ for which $D[\mathfrak{z}] = a$. But by assigning type numbers to *blocs* of minimal surfaces, the usual Morse relations hold.

[1] [1] Shiffman, "The Plateau Problem for Non-Relative Minima," *Ann. Math.*, **40**, 834–854 (1940); [2] Morse and Tompkins, "The Existence of Minimal Surfaces of General Critical Type," *Ibid.*, **40**, 443–472 (1940); [3] Courant, "Critical Points and Unstable Minimal Surfaces," these PROCEEDINGS, **27**, 51–57 (1941); [4] Morse and Tompkins, "Minimal Surfaces Not of Minimum Type by a New Mode of Approximation," *Ann. Math.*, **42**, 62–72 (1941); and "The Continuity of the Area of Harmonic Surfaces as a Function of the Boundary Representations," *Am. Jour. Math.*, **63**, 825–838 (1941).

[2] Cf. [3].

[3] Cf. [4].

[4] See Radò, "On the Problem of Plateau," *Ergeb. Math.*, **2**, 45–47 (1933).

[5] The uniformity of the convergence may be eliminated since it is a consequence of $L^n(1) \rightarrow L^\infty(1)$.

[6] Cf. Shiffman, "The Plateau Problem for Minimal Surfaces Which Are Relative Minima," *Ann. Math.*, **39**, 311–312 (1938).

[7] This theorem is capable of very wide generalizations.

[8] In $\mathfrak{P}$ this metric is equivalent to the uniform metric.

[9] $\mathfrak{P}_a$ consists of those surfaces $\mathfrak{x}$ of $\mathfrak{P}$ for which $D[\mathfrak{x}] \leq a$.

ON HOMOGENEOUS MEASURE ALGEBRAS

BY DOROTHY MAHARAM

THE INSTITUTE FOR ADVANCED STUDY

Communicated January 26, 1942

1. The purpose of this note is to determine the structure of general measure algebras.

2. For any Boolean σ-algebra M, let $\overline{M}$ be the least cardinal number which is the power of a σ-basis of M. M is *homogeneous*, if $\overline{L} = \overline{M}$ for every (principal) ideal $L \subseteq M$ different from the null ideal. A Boolean σ-algebra $M = \{0 \leqq a, b, c, \ldots \leqq e\}$ is a *measure algebra*, if there is defined on M a *measure function* (that is, a countably additive real non-negative function) $\mu(a)$ such that (i) $0 < \mu(e) < \infty$, (ii) $\mu(a) = 0$ if and only if $a = 0$. We also assume that there is no atomic element in M, i.e., (iii) $a > 0$ implies the existence of a $b \in M$ such that $0 < b < a$

An example of a homogeneous measure algebra is the Boolean algebra $P(\gamma)$ of all measurable sets (mod. null sets) of an infinite product space $\Omega_\gamma = P_{0 \leq \alpha < \gamma} I_\alpha$ of intervals I_α: $0 \leqq x_\alpha \leqq 1$, where α and γ are ordinal numbers and the measure on Ω_γ is defined multiplicatively in terms of the ordinary Lebesgue measure on each I_α.

If $\gamma < \gamma'$, then $P(\gamma)$ may be considered as a σ-subalgebra of $P(\gamma')$. This embedding is obtained by identifying each subset E of Ω_γ with the cylinder set $E \times P_{\gamma \leq \alpha < \gamma'} I_\alpha$ in the product space $\Omega_{\gamma'} = \Omega_\gamma \times P_{\gamma \leq \alpha < \gamma'} I_\alpha$.

Two measure algebras are *isomorphic*, if there exists a measure-preserving σ-isomorphism between them. It is easy to see that, for any infinite ordinals γ and γ', two measure algebras $P(\gamma)$ and $P(\gamma')$ are isomorphic if and only if γ and γ' correspond to the same cardinal. Moreover, if γ is finite or corresponds to $\aleph_0$, then $P(\gamma)$ is isomorphic to $P(1)$, i.e., to the measure algebra of Lebesgue measurable sets (mod. null sets) of the interval I: $0 \leqq x \leqq 1$.

THEOREM 1. *Every homogeneous measure algebra with $\mu(e) = 1$ is isomorphic to $P(\gamma_0)$, where γ_0 is the least ordinal number corresponding to M.*

This theorem will be proved by transfinite induction. It will be sufficient to prove the following

LEMMA 1: *Let L be a σ-subalgebra of a homogeneous measure algebra M with $\mu(e) = 1$ such that $\overline{L} < \overline{M}$, and assume that L is isomorphic to $P(\gamma)$, where γ is an ordinal corresponding to $\overline{L}$. Then, for any $a \in M$ with $a \bar{\in} L$, there exists a σ-subalgebra L' of M such that $L \subset L'$, $a \in L'$, and L' is isomorphic to $P(\gamma + 1)$, in such a way that this isomorphism is an extension of the given isomorphism of L and $P(\gamma)$.*

The proof of this lemma will be given in section 4.

3. Let us assume the conditions of Lemma 1. We can consider Ω_γ as a representation space of $L = P(\gamma)$. Then, by a theorem of Radon-Nikodym, there exists, for any $a \in M$, a measurable function $\varphi_a(\xi)$ defined on Ω_γ such that $0 \leqq \varphi_a(\xi) \leqq 1$, and $\int_{E_x} \varphi_a(\xi) d\xi = \mu(a \wedge x)$ for any $x \in L$, where E_x is a measurable set of Ω_γ which corresponds to $x \in L$. An element $a \in M$ is called *independent* of L, if we have $\varphi_a(\xi) =$ constant [namely, $= \mu(a)$] almost everywhere (a. e.) on Ω_γ. This is equivalent to saying that we have $\mu(a \wedge x) = \mu(a) \cdot \mu(x)$ for all $x \in L$.

LEMMA 2: *For any measurable function $\chi(\xi)$ defined on Ω_γ with $0 \leqq \chi(\xi) \leqq \varphi_a(\xi)$ a. e. on Ω_γ, there exists a $b \in M$ such that $0 \leqq b \leqq a$ and $\varphi_b(\xi) = \chi(\xi)$ a. e. on Ω_γ.*

Proof of Lemma 2.—Our lemma is clear if $\chi(\xi) = 0$ a. e. on Ω_γ. Hence we may assume that $\text{meas}[\xi : \chi(\xi) > 0] > 0$. Because of the principle of exhaustion, it is sufficient to show that there exists a $b^* \in M$ such that $0 < b^* \leqq a$ and $\varphi_{b^*}(\xi) \leqq \chi(\xi)$ a. e. on Ω_γ.

Let A be the principal ideal generated by a and let $L(a)$ be the σ-subalgebra of M generated by L and a. Since $\overline{L(a)} = \overline{L} < \overline{A} = \overline{M}$ by assumption,

there exists a $b_1 \in M$ such that $b_1 \in A$ and $b_1 \bar{\in} L(a)$. This means that meas$[\xi: 0 < \varphi_{b_1}(\xi) < \varphi_a(\xi)] > 0$. Again, by the principle of exhaustion, we can further find a $b_2 \in M$ such that $0 < b_2 < a$ and $0 < \varphi_{b_2}(\xi) < \varphi_a(\xi)$ a. e. on the set $[\xi: \varphi_a(\xi) > 0]$. Let us denote by c_1 and c_2 the elements of L which correspond to the sets $[\xi: 0 < \varphi_{b_1}(\xi) \leqq 2^{-1}\varphi_a(\xi)]$ and $[\xi: 2^{-1}\varphi_a(\xi) < \varphi_{b_2}(\xi) < \varphi_a(\xi)]$. If we now put $b_3 = (c_1 \wedge b_2) \vee (c_2 \wedge (a - b_2))$, then $0 < b_3 < a$ and we have $0 < \varphi_{b_3}(\xi) \leqq 2^{-1}\varphi_a(\xi)$ a. e. on the set $[\xi: \varphi_a(\xi) > 0]$. By iterating this argument, we can find, for each n, a $b_{n+2} \in M$ such that $0 < b_{n+2} < a$ and $0 < \varphi_{b_{n+2}}(\xi) \leqq 2^{-n}\varphi_a(\xi)$ a. e. on the set $[\xi: \varphi_a(\xi) > 0]$. Take n so large that meas $E^* > 0$, where $E^* = [\xi: 2^{-n}\varphi_a(\xi) \leqq \chi(\xi)]$, and denote by c^* the element of L which corresponds to E^*. If we put $b^* = c^* \wedge b_{n+2}$, then we have $0 < b^* < a$, and $\varphi_{b^*}(\xi) = \varphi_{b_{n+2}}(\xi) \leqq 2^{-n}\varphi_a(\xi) \leqq \chi(\xi)$ a. e. on E^*, and $\varphi_{b^*}(\xi) = 0$ a. e. on $\Omega_\gamma - E^*$. Hence $\varphi_{b^*}(\xi) \leqq \chi(\xi)$ a. e. on Ω_γ and this proves Lemma 2.

4. *Proof of Lemma 1.* Let Δ_n be the set of all finite sequences $\delta = \{\epsilon_1, \ldots, \epsilon_n\}$, where $\epsilon_i = 0$ or 1, $i = 1, \ldots, n$; and let $\Delta = \mathbf{U}_{n=1}^{\infty}\Delta_n$. A countable set $\{a_\delta\}$ $(\delta \in \Delta)$ is a *dyadic decomposition* of a if $a_0 \vee a_1 = a$, $a_0 \wedge a_1 = 0$, and $a_{\delta,0} \vee a_{\delta,1} = a_\delta, a_{\delta,0} \wedge a_{\delta,1} = 0$ for all $\delta \in \Delta$. By Lemma 2, there exists for any $a \in M$, a dyadic decomposition $\{a_\delta\}(\delta \in \Delta)$ of a such that $\varphi_{a_\delta}(\xi) = \min.(\varphi_a(\xi), \epsilon_1/2 + \ldots + \epsilon_n/2^n + 1/2^n) - \min.(\varphi_a(\xi), \epsilon_1/2 + \ldots + \epsilon_n/2^n)$ a. e. on Ω_γ for all $\delta = \{\epsilon_1, \ldots, \epsilon_n\} \in \Delta$. In the same way, there exists a dyadic decomposition $\{a'_\delta\}(\delta \in \Delta)$ of $a' = e - a$ such that $\varphi_{a'_\delta}(\xi) = \max.(\varphi_a(\xi), \epsilon_1/2 + \ldots + \epsilon_n/2^n + 1/2^n) - \max.(\varphi_a(\xi), \epsilon_1/2 + \ldots + \epsilon_n/2^n)$ a. e. on Ω_δ for all $\delta = \{\epsilon_1, \ldots, \epsilon_n\} \in \Delta$. Let us put $b_\delta = a_\delta \vee a'_\delta$ for all $\delta \in \Delta$. Then $\{b_\delta\}(\delta \in \Delta)$ is a dyadic decomposition of the unit element e, and each b_δ is independent of L, since we have clearly $\varphi_{b_\delta}(\xi) = \varphi_{a_\delta}(\xi) + \varphi_{a'_\delta}(\xi) \equiv 1/2^n$ a. e. on Ω for all $\delta \in \Delta_n$, $n = 1, 2, \ldots$.

Now let L' be the σ-subalgebra of M generated by L and $\{b_\delta\}(\delta \in \Delta)$. L' is obtained by completing the finite algebra L^* consisting of all elements of M of the form: $b^{(n)} = \vee_{\delta \in \Delta_n}(b_\delta \wedge c_\delta)$, where $c_\delta \in L$ for all $\delta \in \Delta_n$, $n = 1, 2, \ldots$. It is easy to see that L' is isomorphic to $P(\gamma + 1)$ by an isomorphism which is an extension of the given isomorphism of L and $P(\gamma)$.

Finally, in order to prove that $a \in L'$, consider the set $[\xi: \varphi_a(\xi) \geqq \epsilon_1/2 + \ldots + \epsilon_n/2^n + 1/2^n]$ for each $\delta = \{\epsilon_1, \ldots, \epsilon_n\} \in \Delta$, and let c_δ be the corresponding element of L. If we put $b^{(n)} = \vee_{\delta \in \Delta_n}(b_\delta \wedge c_\delta)$, then we have $b^{(1)} \leqq b^{(2)} \leqq \ldots \leqq b^{(n)} \leqq \ldots \leqq a$ and $\mu(a - b^{(n)}) \leqq 1/2^n$ for $n = 1, 2, \ldots$. Hence we have $a = \vee_{n=1}^{\infty} b^{(n)}$ and consequently $a \in L'$. This proves Lemma 1 and so Theorem 1.

5. A measure algebra M is a *direct sum* of (a finite or countably infinite number of) measure algebras M_n, if each M_n is (isomorphic to) a principal ideal of M and if every element $a \in M$ is uniquely expressed in the form:

$a = \vee_{n=1}^{\infty} a_n$, $a_m \wedge a_n = 0 (m \neq n)$, where $a_n \in M_n$, $n = 1, 2, \ldots$. (This last condition is equivalent to saying that the unit elements e_n of M_n, which are elements of M, satisfy $e_m \wedge e_n = 0$ $(m \neq n)$ and $e = \vee_{n=1}^{\infty} e_n$.)

THEOREM 2. *Every measure algebra is a direct sum of homogeneous measure algebras $M_n (n = 1, 2, \ldots,$ finite or countably infinite).*

The proof of Theorem 2 is easy and will be omitted.

Theorems 1 and 2 indicate the structure of a general measure algebra.

6. The ergodicity of a measure preserving σ-automorphism T of a measure algebra M (onto itself), or that of a group $G = \{T\}$ of such σ-automorphisms, can be defined as usual.

LEMMA 3. *In order that the group of all measure preserving σ-automorphisms of a measure algebra M be ergodic on M, it is necessary and sufficient that M be homogeneous.*

The proof of this lemma is easy and is omitted.

THEOREM 3. *Let M be a measure algebra, and let G be the group of all measure preserving σ-automorphisms of M. Then M is a direct sum of a countable number of invariant principal ideals M_n, on each of which G is ergodic. This decomposition is the same as in Theorem 2.*

Let G be an arbitrary group of measure preserving σ-automorphisms of M. Then the set L_G of all $a \in M$ such that $T(a) = a$ for all $T \in G$, is a σ-subalgebra of M. Conversely, for any σ-subalgebra $L \subseteq M$, consider the set G_L of all measure preserving σ-automorphisms T of M such that $T(a) = a$ for all $a \in L$. G_L is clearly a group. It is also clear that $L \subseteq L_{G_L}$ and $G \subseteq G_{L_G}$. Exactly as in the theory of Galois, we may ask the question: Under what conditions do the equalities $L = L_{G_L}$ and $G = G_{L_G}$ hold? These and related problems will be discussed elsewhere.

I am greatly indebted to Shizuo Kakutani, who revised and improved my original manuscript.

ON THE DEPOLARIZATION OF NEUTRON BEAMS BY MAGNETIC FIELDS

BY O. HALPERN AND T. HOLSTEIN

RADIATION LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY AND WESTINGHOUSE RESEARCH LABORATORY

Communicated January 10, 1941

In two previous notes[1] referred to henceforth as I and II the authors have discussed the transmission of neutrons through magnetized and unmagnetized ferromagnets and have in detail reported on the physical consequences derivable from observations on polarized neutron beams. For all the applications reference should be made to II.

The mathematical method used was semi-classical; the positional coordinates $\mathbf{r}$ of the neutron were treated fully classically because the wavelength is very small compared to the linear dimensions of the domains in which the magnetic field changes; the behavior of the spin was described with the aid of the self-explanatory equation

$$\frac{d\mathbf{s}}{dt} = g[\mathbf{s} \times \mathbf{B}(\mathbf{r})] \tag{1}$$

which could be integrated for all physically interesting cases as far as the dependence of $\mathbf{B}$ on $\mathbf{r}$ was concerned; the result of this integration, when averaged over all neutron paths, gave directly the change in the *average spin*.

There exist problems of considerable physical interest in which it is necessary to know not only the average spin but the average number of neutrons in each individual spin state. For this type of problem the method described and applied below seems to be appropriate. We shall limit the discussion to two cases which are (cf. II) of direct physical significance. The method is general and applies to all kinds of particles and all values of the spin.

The change in the number of neutrons occupying each spin state is determined as follows: Every individual particle undergoes changes in its spin state under the influence of the magnetic fields. The final result is obtained by averaging over all paths of the various particles which in general will traverse regions of different magnetic field. It will be seen below that the effect of the transmission is to *equalize* the average number of neutrons in each spin state; since a polarized beam is characterized by a preferential occupation of the different spin states, the effect of a ferromagnetic medium is *depolarization* of such a beam.

(1) The magnetic field $\mathbf{B}_i$ of the ith domain which is traversed by a neutron during the time interval τ_i is random in direction but constant in

magnitude and direction during τ_i. The times τ_i are in addition limited by the condition $gB_i\tau_i << 1$. This case constitutes an excellent approximation to the behavior of an unmagnetized polycrystalline specimen, whose crystal grains have linear dimensions $<< 10^{-3}$ cm.

The starting point is the wave equation

$$i\hbar \frac{dc_n}{dt} = \sum_m H_{nm}c_m. \tag{2}$$

Here c_m is the amplitude of the state for which $s_z = n\hbar$ and the H_{nm} are the matrix elements of the Hamiltonian $g\mathbf{s}\cdot\mathbf{B}$ viz.:

$$H_{nn} = n\hbar gB_z \tag{3a}$$

$$H_{n,\,n\pm 1} = \frac{s}{2} g\hbar(B_x \mp iB_y) \sqrt{(l \mp n)(l \pm n + 1)}. \tag{3b}$$

where

$$l = \frac{|s_z|\text{max}}{\hbar}.$$

As before, the positional coördinates are treated classically:

$$x = x_0 + vt,\ y = y_0,\ z = z_0.$$

The existence of randomly oriented fields is taken into account in the following manner: The total time T is divided into intervals τ_i each of which is characterized by a Hamiltonian matrix $H^{(i)}_{nm}$ the average of which over all neutron paths has the following properties:

$$\overline{H^{(i)}_{nm}} = o,\ \overline{H^{(i)}_{nm}H^{(j)}_{mp}} = \delta_{ij}\delta_{np}. \tag{4}$$

The condition $gB_i\tau_i << 1$ becomes $\dfrac{H^i_{nm}\tau_i}{\hbar} << 1$.

One could now find exact solutions of (2), and average by (4); this procedure is somewhat cumbersome and can be avoided by the following method. One introduces a time interval ϑ intermediate in magnitude between the τ_i and T such that: (*a*) many domains are traversed in ϑ; (*b*) the relative change suffered by c_n during ϑ is small. Using the second property, one can immediately write down the values of the c_n at a time $t + \vartheta$ by a method of successive approximations:

$$c_n(t + \vartheta) = c^0_n(t + \vartheta) + c^{(1)}_n(t + \vartheta) + c^{(2)}_n(t + \vartheta) \tag{5}$$

where

$$c_n^0(t+\vartheta) = c_n(t) \tag{5a}$$

$$c_n^{(1)}(t+\vartheta) = \frac{1}{i\hbar}\sum_i\sum_m H_{nm}^{(i)}c_m(t)\tau_i \tag{5b}$$

$$c_n^{(2)}(t+\vartheta) = -\frac{1}{\hbar^2}\sum_i\sum_{j<i}\sum_m\sum_p H_{nm}^{(i)}H_{mp}^{j}c_p(t)\tau_i\tau_j - \frac{2}{\hbar^2}\sum_i\sum_m\sum_p H_{nm}^{(i)}H_{mp}^{(i)}c_p(t)\tau_i^2. \tag{5c}$$

$\sum_i$ here indicates summation over all intervals τ_i which are enclosed in ϑ. The first term of (5c) represents the effect in domains (i) of the changes in c_n produced in previously traversed domains; the second term gives the second order effect of the ith domain, itself. For the probability $w_n(t+\vartheta) = c_nc_n^*$ one obtains, to the second order:

$$w_n(t+\vartheta) = w_n(t) + [c_n^{0*}(t)c_n^{(1)}(t+\vartheta) + \text{comp. conj.}] - \frac{1}{\hbar^2}\sum_{i,i'}\sum_{m,m'} H_{nm}^{i}H_{nm'}^{(i')}c_m(t)c_{m'}(t)\tau_i\tau_i' -$$

$$\frac{1}{\hbar^2}\left[\sum_i\sum_{m,p} H_{nm}^{(i)}H_{mp}^{(i)}c_p(t)c_n^*(t) + \text{compd. conj.}\right] -$$

$$\frac{1}{2H^2}\left[\sum_{i,j<i}\sum_{m,p} H_{nm}^{(i)}H_{mp}^{(j)}c_p(t)c_n^*(t)\tau_i\tau_j + \text{compl. conj.}\right]. \tag{6}$$

For the average over all neutron paths, i.e, for the average over all directions of the magnetic field, one obtains by (4)

$$w_n(t+\vartheta) = w_n(t) + \frac{1}{\hbar^2}\sum_{i,m,m'} H_{nm}^{(i)}H_{nm'}^{(i)}c_m(t)c_{m'}^*(t)\tau_i^2 -$$

$$\frac{1}{2\hbar^2}\left[\sum_{i\,j}\sum_{mp} H_{nm}^{(i)}H_{mp}^{(j)}c_p(t)c_n^*(t)\tau_i\tau_j + \text{compl. conj.}\right] =$$

$$w_n(t) + \frac{1}{\hbar^2}\sum_{im}\overline{|H_{nm}^{(i)}|^2}w_m(t)\tau_i^2 - \frac{1}{\hbar^2}\sum_{i,m}\overline{|H_{nm}^{(i)}|^2}w_n(t)\tau_i^2. \tag{7}$$

In obtaining the last term, the Hermitian property $H_{nm} = H_{mn}^*$ was employed.

Since ϑ is so chosen that only small changes occur in c_n and, consequently in w_n one can replace the differences by differentials, obtaining:

$$\frac{\partial w_n}{\partial t} = \frac{1}{\hbar^2}\frac{\overline{\tau^2}}{\tau_i}\sum_m \overline{|H_{nm}|^2}(w_m - w_n) \cong \frac{\tau}{\hbar^2}\sum_m \overline{|H_{nm}|^2}(w_m - w_n). \quad (8)$$

This equation is an example of a type commonly encountered in statistics. The change in the probability of occupation of a state, n, is equal to the probability of transitions from all other states, m, minus the probability of reverse transitions—from n to m. A stationary condition is reached when all the states are equally occupied. The manner in which equilibrium is attained is somewhat complicated for the general case, but can be illustrated very simply for the particular case of spin $1/2$. Equation (3) together with the equality of $\overline{B_x^2}$, $\overline{B_y^2}\,\overline{B_z^2}$ leads to the equations:

$$\frac{\partial w_{1/2}}{\partial t} = \frac{1}{6}g^2\tau\overline{B^2}(w_{-1/2} - w_{1/2}) = -\frac{\partial w_{-1/2}}{\partial t}. \quad (9)$$

The connection with the classical method can be established by calculating the expectation value for the polarization. In the simple case of spin $1/2$ this polarization s_z is $s/2(w_{1/2} - w_{-1/2})$; the equation for its variation is easily obtained from (9):

$$\frac{\partial s_z}{\partial t} = -\frac{1}{3}g^2\tau\overline{B_z^2}s_z \quad (10)$$

which has the solution

$$s_z = s_{z0}\exp\left(-\frac{1}{3}g^2\tau\overline{B_z^2}t\right) \quad (10a)$$

to be compared with II.

The equation for the spin can readily be obtained in the general case of spin l. Through use of (3) equation (7) becomes

$$\frac{\partial w_n}{\partial t} = \frac{1}{4}\tau g^2(B_x^2 + B_y^2)[(l-n)(l+n+1)(w_{n+1} - w_n) + (l+n)(l-n+1)(w_{n-1} - w_n)]. \quad (11)$$

The equation for the z component of the spin is:

$$\frac{\partial s_z}{\partial t} = \frac{\partial}{\partial t}\sum_n nw_n = \frac{1}{4}g^2\tau(B_x^2 + B_y^2)\sum_n [nw_{n+1}(l-n)(l+n+1) + nw_{n-1}(l+n)(l-n+1) - nw_n\{(l-n)(l+n+1) + (l+n)(l-n+1)\}]. \quad (12)$$

Now

$$\sum_n n w_{n+1}(l-1)(l+n+1) = \sum_{n'=-l+1}^{n'=l+1} (n'-1)(l-n'+1)w_{n'} = \sum_{n=-l}^{n=l} (n-1)(l-n+1)(l+n)w_n$$

since the terms not common to both summations actually have the value zero, as can be learned from inspection. Similarly

$$\sum_{n=-l}^{n=l} n w_{n-1}(l+n)(l-n+1) = \sum_{n=-l}^{n+l} (n+1)(l-n)(l+n+1)w_n.$$

Substituting in (12) one obtains

$$\frac{\partial s_z}{\partial t} = -\frac{1}{2}\tau g^2(\overline{B_x^2} + \overline{B_y^2})\sum n w_n = -\frac{1}{3}\tau g^2 \overline{B^2} s_z \qquad (13)$$

which has the solution

$$s_z(t) = s_{z0} \exp\left(-\frac{1}{3}\tau g^2 \overline{B^2} t\right) \qquad (13a)$$

to be compared with (10*a*).

This formula constitutes the quantum-mechanical verification of the important fact that *the depolarization depends, not on the magnetic moment or the spin separately, but only on their ratio g.*

(2). The total field **B** of each domain is the sum of two fields $\mathbf{B}_0 + \mathbf{B}_i$. The first, $\mathbf{B}_0$, is constant throughout the time T. The second $\mathbf{B}_i$ is constant within τ_i. The direction of $\mathbf{B}_i$ is perpendicular to that of $\mathbf{B}_0$ but otherwise random while the absolute value of B_i is small compared to that of B_0. It will furthermore be assumed that the initial direction of the spin is parallel or anti-parallel to $\mathbf{B}_0$. These assumptions (cf. II) approximate the conditions prevailing during the passage of neutrons through a *quasi saturated polycrystal.*

These conditions take on the following mathematical form:

$$B_{zi} = 0, \overline{B_{xi}B_{yj}} = 0 \text{ unless } i = j, x = y$$

$$\overline{B_{xi}} = \overline{B_{yi}} = 0; \ \overline{B_{xi}^2} = \overline{B_{yi}^2} = \frac{1}{2}\overline{B_i^2}.$$

Due to the fact that $gB\tau_i$ is not small, an approach analogous to that of the preceding section is not applicable. In order to solve (2) explicitly, one sets up a time-dependent perturbation treatment. Introducing the quantities $a_n = c_n \exp(-iE_n t/\hbar)$ where $E_n = H_{nn}$ one obtains the well-known equations

$$i\hbar \frac{da_n}{dt} = \sum_m H_{nm} a_m \exp[-i(E_m - \quad_n)t/\hbar]. \tag{14}$$

If the axis of quantization is chosen parallel to the vector $\mathbf{B}_0$

$$H_{nn} = E_n = ng\hbar B_0 \tag{15a}$$

$$H^{(i)}_{n,\,n\,\pm 1} = {}^1/_2\hbar g(B_{xi} \mp iB_{yi}) \sqrt{(l \mp n)(l \pm n + 1)} \tag{15b}$$

from which it is apparent that the usual mechanism of the time dependent perturbation theory is an expansion in the parameter $|B_i|/|B_0|$.

The solution of (14) is similarly to (5):

$$a_n(t + \vartheta) = a_n(t) + a_n^{(1)}(t + \vartheta) + a_n^{(2)}(t + \vartheta) \tag{16}$$

$$a_n^{(1)}(t + \vartheta) = \sum_i \sum_m H^{(i)}_{nm} a_m(t) \exp(-i(E_m - E_n)\vartheta_i/\hbar)$$

$$\frac{\exp(-i(E_n - E_m)\tau_i/\hbar) - 1}{E_n - E_m} \tag{16a}$$

$$a_n^{(2)}(t + \vartheta) = \sum_i \sum_{m,\,p} H^{(i)}_{nm} H^{(i)}_{mp} a_p(t) \frac{\exp(-i(E_p - E_n)\vartheta/\hbar)}{E_p - E_m}$$

$$\left[\frac{\exp(-i(E_p - E_n)\tau_i/\hbar) - 1}{E_p - E_n} - \frac{\exp(-i(E_m - E_n)\tau_i/\hbar) - 1}{E_m - E_n}\right] +$$

$$\sum_{i,\,j<i} \sum_{m,\,p} H^{(i)}_{nm} H^{(j)}_{mp} a_p(t) \exp[-i/\hbar\{(E_m - E_n)\vartheta_i - (E_p - E_n)\vartheta_j\}]$$

$$\frac{\exp(-i/\hbar(E_n - E_m)\tau_i) - 1}{E_n - E_m} \cdot \frac{\exp(-i/\hbar(E_m - E_p)\,\tau_j) - 1}{E_m - E_p} \tag{16b}$$

with $\vartheta_i = \sum_{j=1}^{j=i-1} \tau_j$.

One now sets up the probability, $w_n(t)$, as before, and averages over the direction of the magnetic fields $\mathbf{B}_i$, obtaining

$$w_n(t + \vartheta) = w_n(t) + 4 \sum_{i,\,m} \frac{\overline{|H_{nm}|^2}}{(E_n - E_m)^2} (w_m - w_n) \sin^2(E_n - E_m)\tau_i/2\hbar. \tag{17a}$$

Replacement of finite differences by differentials yields, finally:

$$\frac{dw_n}{dt} = \sum_m \frac{\overline{|H_{nm}|^2}}{(E_n - E_m)^2} \frac{4}{\tau} (w_m - w_n) \sin^2(E_n - E_m)\tau_i/2\hbar. \tag{17b}$$

Equation (17b) can be integrated without particular difficulty for any case of actual interest. One can on the other hand learn from it without

calculation how the two qualitatively different results of II were arrived at. We found that for $gB\tau << 1$ the depolarization coefficient is proportional to B_i^2 independent of B_0 while in the other case it turned out to be proportional to B_i^2/B_0^2. These results are here immediately obtained from (17*b*) since it is legitimate in the first case to expand the $\sin^2$ thereby removing the energy denominator $(E_n - E_m)^2$ which is proportional to B_0^2. In the second case $\sin^2$ may be replaced by 1/2 leaving thereby the energy denominator $\sim B_0^{-2}$.

These mathematical features permit an informative physical interpretation. In each time interval τ_i matrix elements $\sim gB_i\hbar$ cause transitions between energy levels which are separated by $\sim g\hbar B_0$. The uncertainty principle $\Delta E\tau \sim \hbar$ when applied to this case states that transitions occur with appreciable frequency only between such states which permit conservation to within $\hbar/\tau$. If therefore $gB_0\hbar > \hbar/\tau$ then transitions are cut down; which leads mathematically to a polarization coefficient proportional to B_i^2/B_0^2. If, however, $gB_0\hbar << \hbar/\tau$, then no restrictions on the frequency of transitions are imposed by the energy principle since the quantum theoretical uncertainty exceeds the energy differences between the various states. The frequency of transitions is then solely determined by the values of the non-diagonal matrix elements which are proportional to B_i and independent of B_0.

[1] Halpern, O., and Holstein, T., (I) *Phys. Rev.*, **55**, 601 (1939); (II) *Ibid.*, **59**, 960 (1941).